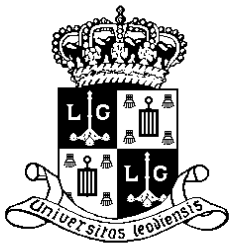


UNIVERSITÉ DE LIÈGE
FACULTÉ DES SCIENCES APPLIQUÉES

Nulling Interferometry with IRSI-DARWIN: Detection and Characterization of Earth-like Exoplanets

Travail de fin d'études présenté par

Olivier ABSIL

en vue de l'obtention du grade
d'ingénieur civil physicien.

Année académique 2000-2001.

Acknowledgements

First of all, I wish to express my very special thanks to Professor Jean Surdej who, in April 2000, drew my attention to the interest and the delight of studying interferometry. I was particularly honoured when he suggested I attend the Geneva meeting on exoplanets and the VLTI. Since that time and throughout the year, he has unceasingly helped and advised me. I am very grateful for his incomparable support.

I would also like to thank Messrs Anders Karlsson and Malcolm Fridlund for giving me the opportunity to work on the DARWIN project. I am particularly grateful to Mr Karlsson for his countless advice and for revising my draft. It was a pleasure to collaborate with him during this year.

My thanks also go to Professors Jean-Pierre Swings, Claude Jamar and Pierre Beckers for accepting to be part of my jury and for examining this work.

I offer my warmest thanks to the "Meudon interferometry team" for welcoming me several times in Meudon, and particularly to Bertrand and Vincent for their precious advice. A special thank also to Professor Pierre Léna for constructive discussions about choosing the subject of this thesis, and for his enthusiasm and support.

I am also grateful to Mr Pierre Rochus for registering me to the "Journées Scientifiques de l'ONERA", for guiding me in Paris up to Meudon, and for handing me several valuable documents.

I take a particular pleasure in thanking all members of my family for their moral support and for bearing my changing moods these last weeks. A very special thank to Pierre-Antoine for fruitful discussions on signal-to-noise calculations, and for taking care of the network. A big thank also to André and Mamy for correcting my English and giving another point of view on this work.

Finally, I would like to thank my own Lapinou wholeheartedly for enduring my long dissertations on interferometry, for raising my spirits when needed, and for giving me so much love.

Contents

Preface

There is an infinite number of worlds, both like and unlike this world of ours.
Epicurus, Letter to Herodotus (341–270 BC)

There are countless suns and countless earths all rotating around their sun in exactly the same way as the seven planets of our system. We see only the suns because they are the largest bodies and are luminous, but their planets remain invisible to us because they are smaller and non-luminous. The countless worlds in the universe are no worse and no less inhabited than our Earth.
Giordano Bruno, De L'Infinito Universo E Mondi (1584)

I should not hesitate to stake my all on the truth of the proposition—if there were any possibility of bringing it to the test of experience—that at least some one of the planets which we see is inhabited. Hence I say that I have not merely the opinion, but the strong belief that there are inhabitants in other worlds.
Immanuel Kant, The Critique of Pure Reason (1781)

It would indeed be rash to assume that nowhere else in the Universe has nature repeated the strange experiment which she has performed on the Earth.
Arthur Eddington, Nature of the Physical World (1933)

There are hundreds of billions of galaxies in the observable Universe, with each galaxy such as our own containing 10^{11} stars. Surrounded by this seemingly limitless ocean of stars, mankind has long speculated about the existence of planetary systems other than our own, and the possibility of development of life elsewhere in the Universe. We live in a remarkable time, where human beings have finally attained the possibility of finding the first real answers to some of these most meaningful questions.

Since the first discovery of an exoplanet[1] by Mayor and Queloz in 1995, the hunt for such bodies has gained much interest in the scientific community, to such a point that more than 60 exoplanets have already been detected and confirmed to date. Currently, about one exoplanet is found every week. But this ground-based hunt for extra-solar planets does not have the sensitivity required to detect planets similar to our Earth: all 60 exoplanets are as

[1]The neologism "exoplanet" refers to a planet orbiting around another star than our own.

big as Jupiter, and generally orbiting very close to their parent star, so that their surface temperature is quite high (900 K or hotter). These planets are nicknamed "Hot Jupiters". Moreover, the present detection techniques provide very little information about the planet itself (essentially its mass and revolution period) because it is only detected through its gravitational influence on its host star: present telescopes do not allow to directly see the planet. This is going to change thanks to space interferometry, which is the main subject of this work.

In the first chapter of this thesis, a brief introduction to the principles of interferometry is given. After describing Young's experiment, which shows how two light beams interfere, we recall the basic equations of interferometry by means of some analytical developments. A short historical review of stellar interferometry is then given, followed with some practical considerations, including the effect of the turbulent atmosphere on ground-based interferometers. The last paragraphs of that chapter put the emphasis on technical difficulties faced by stellar interferometry, and show how these difficulties are or will be overcome. The chapter is targeted at readers unfamiliar with stellar interferometry. It should help understand the following chapters more easily, and is a necessary complement to chapter 5.

The second chapter first deals with the various techniques currently used for exoplanet detection. Since planets are typically billions of times fainter than their host star, with an angular separation of at most a few seconds of arc, direct detection is an extraordinarily difficult task. This is the reason why *indirect detection* methods, based on dynamical perturbation of the star by the orbiting planet, on planetary transits and on gravitational microlensing, have been developed. The second part of chapter 2 is dedicated to *direct detection* methods, which will be available in a near- or mid-term future. Finally, basic considerations on remote sensing of life are used to show why the mid-infrared range is the most interesting for a mission dedicated to planet and life finding.

The third chapter consists in a detailed presentation of the IRSI-DARWIN[2] mission concept, with particular attention to the interferometric configuration (or "aperture" configuration). DARWIN is a candidate Cornerstone to the Horizon 2000+ program of the European Space Agency (ESA). It would consist of half-a-dozen small telescopes, each like a small Hubble telescope, separated by about fifty meters, and sending their light to a central station, in a technique known as *nulling interferometry*. We show how the DARWIN concept has evolved since its first proposal in 1993, and how it will address the three questions of detecting exoplanets, characterizing their habitability, and observing signposts for the existence of life as we know it. A very important breakthrough for the DARWIN mission is the development of *internal modulation*, a technique for fast signal chopping described in section 3.2.5. Some examples of configurations with internal modulation are then presented, as well as the configurations proposed for the Terrestrial Planet Finder (the NASA equivalent of DARWIN). This chapter ends with a proof of the feasibility of the DARWIN mission, based on signal-to-noise calculations.

The intended contribution of this thesis to the DARWIN mission consists in a systematic study of possible aperture configurations with internal modulation. Chapter 4 is dedicated to that study. For technological reasons, the telescopes forming the nulling interferometer

[2]Infra-Red Space Interferometer DARWIN.

are restricted to stand on a circle or on a common line. Even with this restriction, there are countless possibilities for the DARWIN aperture configurations. Their performances are evaluated by obtaining analytical expressions for the *transmission map* and the *modulation efficiency* of each interferometer. A new way to modulate the signal, called *inherent* internal modulation, is proposed in section 4.4. This new kind of internal modulation is especially efficient with linear arrays, because in that case it uses fewer detectors than classical internal modulation, with the same performance. Criteria for the choice of an optimal configuration are discussed at the end of the chapter.

One of the key technology developments needed for DARWIN is the design and test of nulling interferometry. The installation of a nulling breadboard on ESO's VLT interferometer not only could validate the technique of nulling interferometry but also could provide much-needed scientific information about the prevalence of dust clouds in nearby planetary systems. Chapter 5 deals with this possibility. After a brief description of zodiacal dust clouds, an evaluation of three possible nulling configurations is undertaken, with the result that a two-telescope interferometer is probably the most appropriate. The main technical difficulties associated with ground-based nulling interferometry are then briefly described, as are other current projects.

Four appendices contain analytical calculations that are not critical for a good understanding of the work, but, on the other hand, are fundamental for the systematic study of aperture configurations. These calculations are mainly based on the properties of linear and nonlinear systems of equations, and can be skipped at first reading.

Olivier Absil
Liège, May 2001.

Notations and acronyms

Conventions

$\vec{A}$:	vector
$\mathbf{A}$:	complex variable or function
$\overline{\mathbf{A}}$:	complex conjugate of $\mathbf{A}$
$\widetilde{\mathbf{A}}$:	Fourier transform of $\mathbf{A}$
$[\pi]$:	to within an integer multiple of π

Units

arcsec:	second of arc ($\simeq 4.85 \times 10^{-6}$ radian)
mas:	milli-arcsec (10^{-3} arcsec)
μas:	micro-arcsec (10^{-6} arcsec)
AU:	Astronomical Unit ($\simeq 1.495 \times 10^{8}$ km)
pc:	parsec ($\simeq 3.08 \times 10^{14}$ km)
Jy:	Jansky (10^{-26} W m^{-2} Hz^{-1})

Constants

$M_\odot$:	Sun's mass ($\simeq 1.98 \times 10^{30}$ kg)
$M_\oplus$:	Earth's mass ($\simeq 5.98 \times 10^{24}$ kg)
M_J:	Jupiter's mass ($\simeq 1.9 \times 10^{27}$ kg)
G:	Cavendish gravitational constant ($\simeq 6.67 \times 10^{-11}$ m^{3}kg^{-1}s^{-2})

Notations

M_*, R_*, L_*:	mass, radius and luminosity of the star
M_p, R_p, L_p:	mass, radius and luminosity of the exoplanet
a:	semi-major axis of the exo-system
P:	orbital period of the exo-system
i:	inclination of the exoplanet's orbital plane with respect to the plane of the sky
B_ν, B_λ:	surface brightness, respectively per frequency and per wavelength unit
F_ν, F_λ:	flux density, respectively per frequency and per wavelength unit

$\vec{\theta} = (\theta, \phi)$:	angular coordinates of a source in the plane of the sky
$\vec{r}$:	cartesian coordinates in the pupil plane
$\vec{\theta_f}$:	angular coordinates in the focal plane
$R(\theta, \phi)$:	interferometer response
$\vec{L_k} = (L_k, \delta_k)$:	polar coordinates of telescope k with respect to the origin of the frame (generally taken as the center of the array)
D_k:	diameter of telescope k
ϕ_k:	phase shift applied to the beam from telescope k
c_k:	$c_k = D_k \cos \phi_k$, amplitude of telescope k in a real entrance pupil
s_k:	$s_k = D_k \sin \phi_k$, non-zero only when the entrance pupil is complex

Acronyms

AIC:	Achromatic Interfero Coronagraph
ALMA:	Atacama Large Millimeter Array
COBE:	Cosmic Background Explorer
DAC:	Degenerated Angel Cross
ESA:	European Space Agency
ESO:	European Southern Observatory
FTRC:	Five Telescope Real Circle
GAC:	Generalized Angel Cross
HST:	Hubble Space Telescope
IR:	Infrared
IRAS:	Infra-Red Astronomical Satellite
IRSI-DARWIN:	Infra-Red Space Interferometer DARWIN
LBT:	Large Binocular Telescope
MMT:	Multiple Mirror Telescope
NGST:	Next Generation Space Telescope
OASES:	Outpost for Analysis and Spectroscopy of Exo-Systems
OPD:	Optical Path Difference
PSF:	Point Spread Function
RL$x(m_1,\ldots,m_n)$:	Robin Laurance configuration, where x is the number of sub-arrays, n the number of telescopes and m_i the relative telescope size.
rms:	root mean square
SIM:	Space Interferometry Mission
SIRTF:	Space Infra-Red Telescope Facility
SNR:	Signal-to-Noise Ratio
SOFIA:	Stratospheric Observatory for Infrared Astronomy
STRC:	Six Telescope Real Circle
TPF:	Terrestrial Planet Finder
VLTI:	Very Large Telescope Interferometer

Chapter 1

Principles of optical/IR interferometry

1.1 Coherence of light

The principles of interferometry are profoundly based on the theory of light coherence. By re-visiting Young's experiment, we will introduce the concepts of temporal and spatial coherence, as well as the basic equations of interferometry. This chapter is largely based on Born and Wolf [14], Léna [33], Surdej [53] and on the course notes from the 1999 Michelson Summer School [28].

1.1.1 Young's experiment

The earliest experimental setting for demonstrating the interference of light is due to Young. Light from a monochromatic point source S falls on two pinholes S_1 and S_2 which are close together in a screen $\mathcal{A}$ and equidistant from S (figure 1.1). The pinholes act as secondary monochromatic point sources which are in phase, and the beams from them are superposed in the region beyond $\mathcal{A}$. An interference pattern is formed on screen $\mathcal{B}$ placed at a distance d from $\mathcal{A}$.

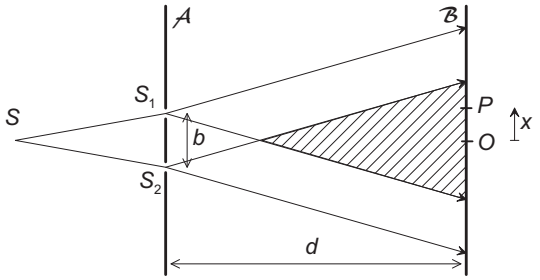


Figure 1.1: Young's experiment: light from a monochromatic point source S is divided into two spherical wavefronts by two pinholes S_1 and S_2, separated by a distance b called the baseline. These two wavefronts create an interference pattern on screen $\mathcal{B}$.

By expressing the optical path difference (OPD) for light reaching point P (with coordinate x on screen $\mathcal{B}$) from S_1 and S_2, and remembering that there are respectively intensity

maxima and minima when the two waves are in phase or half a wavelength out of phase, one finds the following positions for intensity extrema:

$$x_{max} = m\frac{d\lambda_0}{b} \quad \text{or} \quad x_{min} = (2m+1)\frac{d\lambda_0}{2b}\,, \qquad |m| = 0, 1, 2, \ldots\,, \tag{1.1}$$

where λ_0 is the wavelength and m the order of interference. The interference pattern in the immediate vicinity of O thus consists of bright and dark bands called interference fringes, equidistant and running at right angle to the line S_1S_2 joining the two sources. The separation between adjacent bright fringes is $d\lambda_0/b$.

Now if the source is not perfectly monochromatic, it can be seen as a superposition of mutually incoherent monochromatic components, each producing an interference pattern with a fringe spacing proportional to its wavelength λ. The central maxima of all the monochromatic patterns, corresponding to equality of the paths from S_1 and S_2, coincide on O, where a bright "white" fringe is observed, but elsewhere the patterns are mutually displaced because their scale is proportional to λ. As a result, the fringes are less distinct than with monochromatic light, and the situation gets worse when the wavelength range $\Delta\lambda$ increases.

Another way to reduce the contrast between fringes has been discovered by Fizeau in 1868: the fringe contrast decreases when the angular diameter of the source S increases, or when the baseline b between the two holes increases. Based on this property, the angular distance θ between two close unresolved stars can be determined in a (theoretically) very simple way: when increasing the distance between the two holes, the fringe patterns from the two stars progressively separate until the bright fringes from one pattern exactly coincide with the dark fringes of the other (see figure 1.2). At that very moment the two stars are resolved, and their angular separation is given by $\theta = \lambda_0/2b$.

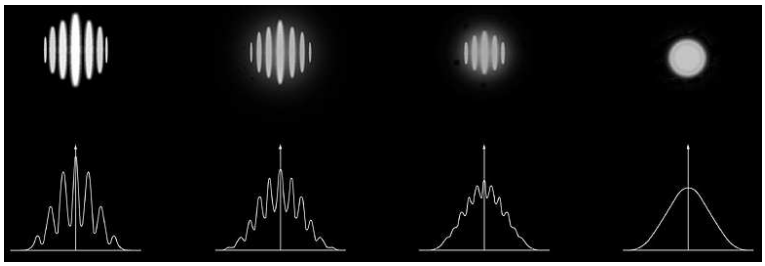


Figure 1.2: Interference fringes for a binary star. From left to right: the contrast between bright and dark fringes progressively decreases as the baseline increases, until the moment where it cancels and the two sources are resolved (ESO PR Photo 10e/01).

1.1.2 Complex degree of coherence and fringe visibility

Consider the wave field $\mathbf{V}(P,t)$ produced by an extended polychromatic source, and two points P_1 and P_2 in the wave field. The observable intensity at these two points is equal to the mean value of $|\mathbf{V}(P_j,t)|^2$. In addition to measuring the intensities $I(P_j) =$

$\langle \mathbf{V}(P_j,t)\overline{\mathbf{V}}(P_j,t)\rangle$ at each of these points, we may also determine experimentally the interference effects arising from the superposition of the vibrations from these points by repeating Young's experiment. By placing an opaque screen $\mathcal{A}$ across the field with pinholes at P_1 and P_2, the field distribution on a screen $\mathcal{B}$ placed some distance from $\mathcal{A}$ reads

$$\mathbf{V}(Q,t) = \mathbf{K}_1\mathbf{V}(P_1,t-t_1) + \mathbf{K}_2\mathbf{V}(P_2,t-t_2) \;\Leftrightarrow\; \mathbf{V}_Q(t) = \mathbf{K}_1\mathbf{V}_1(t) + \mathbf{K}_2\mathbf{V}_2(t-\tau)\;, \tag{1.2}$$

with t_1, t_2 the times needed for light to travel from P_1 or P_2 to Q, and with $\tau = t_2 - t_1$, assuming that the field is stationary. The pure imaginary[1] factors $\mathbf{K}_1$ and $\mathbf{K}_2$ are inversely proportional to the distances P_1Q and P_2Q, and also depend on the size of the openings and on the geometry of the arrangement. The intensity at point Q then writes

$$I(Q) = |\mathbf{K}_1|^2 I_1 + |\mathbf{K}_2|^2 I_2 + 2|\mathbf{K}_1\mathbf{K}_2|\,\Re(\Gamma_{12}(\tau))\;, \tag{1.3}$$

where $\Gamma_{12}(\tau) = \langle \overline{\mathbf{V}}_1(t)\mathbf{V}_2(t-\tau)\rangle$ is the mutual coherence of the light vibrations at P_1 and P_2, the vibrations at P_2 being considered τ seconds earlier than at P_1. Normalizing this function, we obtain the complex degree of coherence

$$\boldsymbol{\gamma}_{12}(\tau) = \frac{\Gamma_{12}(\tau)}{\sqrt{\Gamma_{11}(0)}\sqrt{\Gamma_{22}(0)}} = \frac{\Gamma_{12}(\tau)}{\sqrt{I_1}\sqrt{I_2}} \tag{1.4}$$

which simultaneously describes the spatial and temporal coherence of the waves because it correlates the fields at two different points of space and time.

Considering a quasi-monochromatic source (with bandwidth $\Delta\nu \ll \nu_0$), one defines the coherence time of the source $\tau_c \sim 1/\Delta\nu$ as the characteristic times separating two instants at which the source electric field begins to be decorrelated. Now if $\tau \ll \tau_c$, it can be shown that

$$\boldsymbol{\gamma}_{12}(\tau) = |\boldsymbol{\gamma}_{12}(\tau)|\exp(i\alpha_{12}(\tau) - i2\pi\nu_0\tau) \simeq |\boldsymbol{\gamma}_{12}(0)|\exp(i\alpha_{12}(0) - i2\pi\nu_0\tau)\;, \tag{1.5}$$

so that the intensity at point Q writes[2]

$$I(Q) = I_1(Q) + I_2(Q) + 2\sqrt{I_1(Q)I_2(Q)}|\boldsymbol{\gamma}_{12}(0)|\cos(\alpha_{12}(0) - 2\pi\nu_0\tau)\;. \tag{1.6}$$

Hence the visibility of the fringes at Q is

$$\mathcal{V}(Q) = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = \frac{2\sqrt{I_1(Q)I_2(Q)}}{I_1(Q) + I_2(Q)}|\boldsymbol{\gamma}_{12}(0)|\,, \tag{1.7}$$

which reduces to $\mathcal{V}(Q) = |\boldsymbol{\gamma}_{12}(0)|$ in the frequent case where the two beams are of equal intensity. A direct measurement of $|\boldsymbol{\gamma}_{12}(0)|$ is thus possible by measuring the fringes visibility. Its value depends on the extension of the source and on the distance between P_1 and P_2, according to the van Cittert–Zernicke theorem, discussed in the following section.

[1] Since the secondary wavelets from P_1 and P_2 are out of phase with the primary wave by a quarter of period, $\mathbf{K}_1$ and $\mathbf{K}_2$ are pure imaginary numbers (see [14]).

[2] The intensity $I_j(Q) = |\mathbf{K}_j|^2 I_j$ represents the intensity that would be measured at point Q if there was a unique pinhole at the point P_j in the screen $\mathcal{A}$, and thus no interference.

1.1.3 The van Cittert–Zernicke theorem

In this section, we derive the fundamental theorem of interferometry, which connects the complex degree of coherence $\boldsymbol{\gamma}_{12}(\tau)$ to the intensity distribution of the source.

Let us re-write the expression of $\boldsymbol{\gamma}_{12}(\tau)$ in the case where the extended source S is divided into elements $\mathrm{d}\sigma_1$, $\mathrm{d}\sigma_2$, ... centered on points S_1, S_2, ... of small linear dimensions compared to the mean wavelength λ_0. If $\mathbf{V}_{m1}(t)$ and $\mathbf{V}_{m2}(t)$ are the complex disturbances at P_1 and P_2 due to the source element $\mathrm{d}\sigma_m$, the total disturbances at these points are

$$V_1(t) = \sum_m V_{m1}(t) \;, \qquad V_2(t) = \sum_m V_{m2}(t) \;. \tag{1.8}$$

The mutual intensity thus reads

$$\Gamma_{12}(0) = \langle \overline{\mathbf{V}}_1(t)\mathbf{V}_2(t)\rangle = \sum_m \langle \overline{\mathbf{V}}_{m1}(t)\mathbf{V}_{m2}(t)\rangle + \sum_m \sum_{n\neq m} \langle \overline{\mathbf{V}}_{m1}(t)\mathbf{V}_{n2}(t)\rangle \;. \tag{1.9}$$

Now the light vibrations arising from different elements of the source may be assumed to be statistically independent (mutually incoherent), and of zero mean value so that

$$\langle \overline{\mathbf{V}}_{m1}(t)\mathbf{V}_{n2}(t)\rangle = \langle \overline{\mathbf{V}}_{m1}(t)\rangle\langle \mathbf{V}_{n2}(t)\rangle = 0 \;\; \text{when} \;\; m \neq n \;. \tag{1.10}$$

If R_{m1} and R_{m2} are the distances of P_1 and P_2 from the source element $\mathrm{d}\sigma_m$, then

$$\mathbf{V}_{mj}(t) = \mathbf{A}_m(t - R_{mj}/c)\frac{\exp(i2\pi\nu_0(t - R_{mj}/c))}{R_{mj}} \;, \qquad j = 1, 2 \;, \tag{1.11}$$

where $\mathbf{A}_m$ is the amplitude of the radiation and c the speed of light. Hence

$$\langle \overline{\mathbf{V}}_{m1}(t)\mathbf{V}_{m2}(t)\rangle = \langle \overline{\mathbf{A}}_m(t)\mathbf{A}_m(t - (R_{m2} - R_{m1})/c)\rangle \frac{\exp(i2\pi\nu_0(R_{m1} - R_{m2})/c)}{R_{m1}R_{m2}} \;. \tag{1.12}$$

If the path difference $R_{m2} - R_{m1}$ is small compared to the coherence length $l_c = c\tau_c$ of light we may neglect the retardation $(R_{m2} - R_{m1})/c$ in the argument of $\overline{\mathbf{A}}_m$, and the expression for the mutual intensity reads

$$\Gamma_{12}(0) \simeq \sum_m \langle \overline{\mathbf{A}}_m(t)\mathbf{A}_m(t)\rangle \frac{\exp(i2\pi\nu_0(R_{m1} - R_{m2})/c)}{R_{m1}R_{m2}} \;. \tag{1.13}$$

Finally, denoting by $I(S)$ the intensity per unit area of the source, so that $I(S_m)\mathrm{d}\sigma_m = \langle \overline{\mathbf{A}}_m(t)\mathbf{A}_m(t)\rangle$, the complex degree of coherence writes

$$\boldsymbol{\gamma}_{12}(0) = \frac{1}{\sqrt{I_1 I_2}} \int_\sigma I(S) \frac{\exp\left(i2\pi\frac{R_1 - R_2}{\lambda_0}\right)}{R_1 R_2} \,\mathrm{d}S \;, \tag{1.14}$$

where R_1 and R_2 denote the distances between a typical source point S and points P_1 and P_2 (see figure 1.3), and I_1, I_2 the intensities at P_1 and P_2. This is the mathematical

expression of the van Cittert–Zernicke theorem, which can be simplified in the important case where the distance between the source S and the screen $\mathcal{A}$ is very large (see [14]):

$$\gamma_{12}(0,x,y) = e^{i\psi}\frac{\int\int_\sigma I(\xi,\eta)\exp(-i2\pi(x\xi+y\eta)/\lambda)\mathrm{d}\xi\mathrm{d}\eta}{\int\int_\sigma I(\xi,\eta)\mathrm{d}\xi\mathrm{d}\eta}\ , \tag{1.15}$$

where $x = x_1 - x_2$ and $y = y_1 - y_2$ are the coordinate differences between points P_1 and P_2, and (ξ,η) the coordinates of a typical source point S (see figure 1.3). In this expression, the quantity ψ represents the phase difference $2\pi(OP_1 - OP_2)/\lambda_0$. It is further conventional to define the spatial frequencies $u = x/\lambda$ and $v = y/\lambda$.

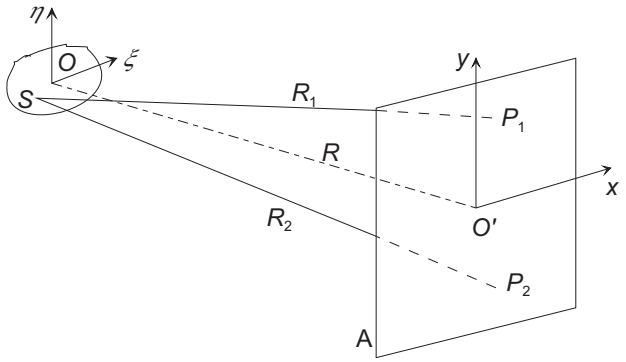


Figure 1.3: Illustrating the van Cittert–Zernicke theorem.

In short, the van Cittert–Zernicke theorem states that the complex degree of mutual coherence (and therefore the fringe visibility) is essentially the normalized two-dimensional Fourier transform of the source intensity. We will see in section 1.2.4 how this theorem can be used in order to image an unresolved source with an interferometer. But first, let us briefly describe how a stellar interferometer works.

1.2 Stellar interferometry

1.2.1 From Young to Fizeau

As already mentioned in section 1.1, Fizeau proposed in 1868 a way to determine the angular size of an unresolved star[3], by progressively increasing the distance b between two holes until the visibility of the fringes reaches a minimum. An experiment based on this property was carried out by Stephan in 1873, who placed an opaque screen with two holes in front of an 80 cm telescope (figure 1.4). With this 65 cm baseline, he did not observe visibility loss for any target star, and concluded that stellar angular diameters are much smaller than 0.16 arcsec.

A more general Fizeau configuration is shown on the right-hand side of figure 1.4. It consists in two mirrors lying on a virtual parabola (dotted curve), so that the two parallel

[3]or equivalently the angular separation between the two components of an unresolved binary star

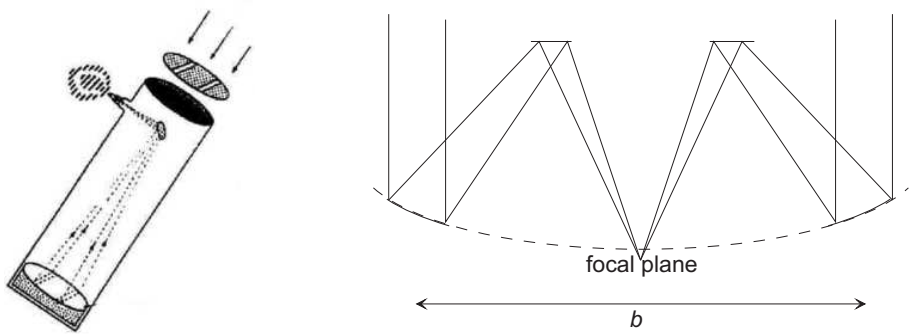


Figure 1.4: Left: Optical layout of a rudimentary Fizeau interferometer, as used by Stephan (1873). Right: Generalization of the Fizeau configuration to a two-telescope interferometer.

light beams are focused at a common point. Secondary mirrors are added in order to improve the compactness of the system. Interferences between the two beams at the focus result as usual in a fringe pattern. Long baselines are still difficult to achieve on ground with such interferometers because the two mirrors have to be mounted on a common support, just as if it was a single mirror. In addition, the structure has to be rotated as a whole in order to point at a non-zenith star.

1.2.2 The Michelson stellar interferometer

In order to increase the interferometer baseline, Michelson installed a six-meter metal beam with small mirrors at each end across the aperture of the 100-inch (2.5-meter) Hooker telescope at Mount Wilson in California, then the biggest in the world (figure 1.5). With this interferometer, Michelson succeeded in measuring the angular diameter of half a dozen of red giants, such as Betelgeuse (0.047 arcsec). Anderson repeated the same experiment in order to resolve very tight binaries such as Capella.

Figure 1.5: Left: Optical layout of a rudimentary Michelson stellar interferometer, as used by Michelson (1919). Right: Photograph of the 2.5-m Hooker telescope surmounted by a 6-m metal beam. Taken from the Space Interferometry Mission (SIM) web-site.

In the 1930s Michelson's collaborator F. Pease attempted interferometry with mirrors on a 15-meter mount, but he failed, probably because the mount was not mechanically stiff enough. Indeed the optical path difference between the two interfering beams must be much smaller than the coherence length of the wave (~ 1 μm in the optical range) in order to record the fringes, and a 15-m beam does not insure such an accuracy.

The next step was naturally to abandon single aperture interferometers, and thus to use two independent telescopes which could be separated by any chosen distance. This step was finally made by Labeyrie in 1973. The optical layout of a "modern" Michelson interferometer is shown in figure 1.6. The two afocal beams coming from the telescopes are recombined in a common focal plane, and the fringe spacing is determined by the distance d between the two parallel beams at recombination.

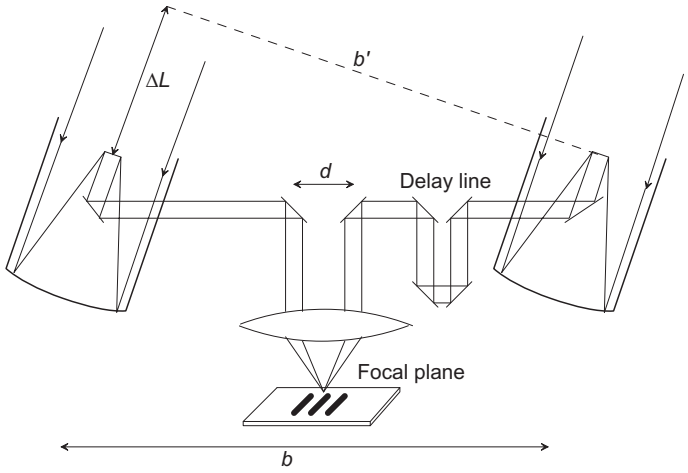


Figure 1.6: Optical layout of a modern Michelson interferometer. A delay line introduces a variable optical delay in order to compensate for the ΔL optical path difference between the two samples of the incoming wavefront. The distance d between the two parallel beams at recombination determines the fringe spacing in the focal plane, where the fringe visibility $|\gamma_{12}(0)|$ is measured.

1.2.3 Image-plane and pupil-plane combination

There are two fundamentally different types of beam-combination at the back end of an interferometer, and all ground- and space-based interferometers use one or the other of these methods. In principle, with an ideal instrument, the best achievable signal-to-noise ratios from both methods should be identical.

Image-plane interferometry

It is the method in which each beam is focused to make an image of the sky, and the images are superposed in such a way that interferences fringes form across the combined image. This is also called multi-axial interferometry, or sometimes "Fizeau interferometry", not to

be confused with the "Fizeau configuration" shown in figure 1.4. Image-plane combination has been implicitly assumed until now (see for example figure 1.6 where a lens is used to focus the two beams on a detector). With the notations of figure 1.6, the fringe intensity as a function of angular position θ in the focal plane is given by

$$I(\theta) = 2I_{tel}(\theta)(1 + \mathcal{V}\cos(2\pi(\theta d + \Delta z)/\lambda)) \,, \tag{1.16}$$

where I_{tel} is the shape of the diffraction envelope, $\mathcal{V}$ the visibility of the star, θd the fringe modulation term and Δz the residual OPD, which determines the fringe position. This residual OPD depends on the baseline b and on the line-of-sight. When the delay line is adjusted to give an OPD of $\Delta z = 0$, then the peak intensity of the fringe pattern is centered in the diffraction envelope (Airy pattern) of the two beams. If the delay line is moved off the zero OPD position, the envelope will stay fixed but the fringes will move across the envelope.

Pupil-plane interferometry

It is the method in which parallel beams are superposed, using a half-silvered mirror or equivalent, and the two resulting output beams are each focused on single detector pixels (figure 1.7). This is also called co-axial interferometry, or sometimes "Michelson interferometry", after the eponym's original 1893 interferometer[4], which showed that the speed of light is independent of the observer velocity. In pupil-plane interferometry, the combining beams are completely overlapped, so the combining baseline is zero ($d = 0$) and the interfringe distance infinite.

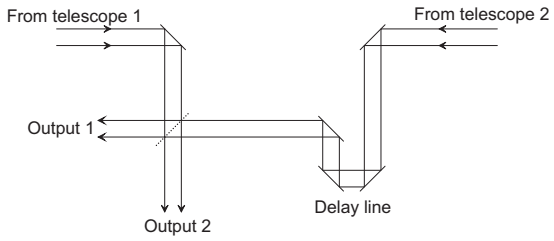


Figure 1.7: Pupil plane combination schematic, showing a half-silvered beam-combiner plate and zero distance between combining beams.

Since a symmetric beam-splitter has the property[5] that the phase difference between transmitted and reflected beams is exactly $\pi/2$, the phase difference between the combined beams is $2\pi\Delta z/\lambda \pm \pi/2$, so that the fringe intensity is modulated by a $\pm\sin$ term instead of a cos:

$$I(t) = 2I_{tel}(1 \pm \mathcal{V}\sin(2\pi\Delta z(t)/\lambda)) \,. \tag{1.17}$$

Here we have integrated over an angle in the focal plane so that $I_{tel} = \int I_{tel}(\theta)\mathrm{d}\theta$, and we have explicitly assumed that the phase difference is time-modulated rather than being spatially modulated as in the image-plane case.

[4] not to be confused with his 1919 stellar interferometer

[5] See Lawson [28], chap. 3, for an analytical proof.

1.2.4 Principles of aperture synthesis

An interferometer is not an imager in and of itself: it only produces fringes, which give information on the object at a single spatial frequency determined by the baseline. But in principle, one may completely and unambiguously reconstruct an astronomical scene at least up to the limiting b/λ frequency by using aperture synthesis techniques developed by radio astronomers to systematically measure in time all the different fringe patterns obtained on the source with a large number of different baseline lengths and orientations.

In practice, a set of discrete interferometer baselines $\vec{b}_i = (x_i, y_i)$ yields a set of discrete visibility measurements $\mathcal{V}(u_i, v_i)$. This discrete visibility field can then be inverted by means of a discrete Fourier transform operation to obtain a bandwidth-limited estimate of the parent brightness distribution. The accuracy of the synthesized image is naturally a function of the coverage of the (u, v) plane. However in order to properly image a source in two dimensions it is necessary to accurately measure not only the contrast $\mathcal{V}(u, v) = |\boldsymbol{\gamma}_{12}(0, u, v)|$ of the fringe packet but also its precise position or phase with respect to some reference point. This is much harder to accomplish accurately due to the randomly varying nature of some sources of OPD in an interferometer. These sources are briefly discussed in section 1.3.2.

1.3 Practical limitations to optical/IR interferometry

A first difficulty that faces practical application of interferometry is related to the inherent nature of the technique that has to rely, as was already apparent in figure 1.2, on the measurement of the visibility of a fringe packet where it is least detectable, namely where it disappears! For complex sources of any type, the visibility is low by definition if the interferometer is resolving the object at all. It can be shown, in fact, that the accuracy with which an interferometric measurement can be made depends on the product $N\mathcal{V}^2$ where N is number of photons collected by an individual telescope in the array and $\mathcal{V}$ is the measured visibility. This implies that interferometry becomes increasingly inefficient as the source is resolved. Fortunately there are several ways around this particular problem, such as wavelength bootstrapping, baseline bootstrapping or guide star methods (see Lawson [28], chap. 14).

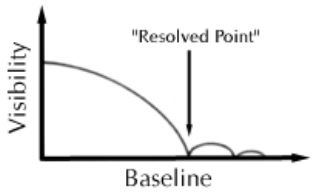


Figure 1.8: Fringe visibility vs. baseline. Measurements of low visibilities are necessary in order to characterize the source, but unfortunately these measurements are the less accurate.

1.3.1 Visibility loss effects

High-quality measurements require that the observer minimize and calibrate the instrumental losses of visibility. Five main visibility loss effects are discussed by Traub in [28], and are briefly summarized hereafter.

- *Spectral bandpass.* All stellar measurements use a finite range of wavelengths. Any fringe packet, whether it is displayed spatially or temporally, will suffer a reduction in modulation amplitude at the edges of the packet, where the different wavelengths will produce opposing peaks and valleys. It can be shown that the number of fringes between envelope zero-crossings in a finite-bandpass wave packet is $N = 2\lambda/\Delta\lambda$.

- *Wavefront tilt.* If two wavefronts of size D are tilted by an angle α, then the interference pattern will be smeared and the visibility reduced. For a two-dimensional circular aperture of diameter D, the visibility reduction is smaller than 10% if $\alpha < 0.3\lambda/D$.

- *Intensity mismatch.* If the relay optics fail to perfectly overlap the beams from each telescope, or if the beam-combiner has unequal reflection and transmission factors, or if the combined beams come from different diameter telescopes and therefore have different intensities in the overlap region, then we will have a reduction in visibility from any of these factors. This is a relatively tolerant effect.

- *Optical surface figure error.* If the combining wavefronts each have a rms perturbation of δ with respect to a perfect wavefront, and if the perturbations are randomly distributed across the wavefront, and uncorrelated between the two wavefronts, then the fringe visibility will be degraded by a factor $\exp(-(2\pi\delta/\lambda)^2)$. This is a significant loss, especially when there is a large number of reflections: the cumulative effect of even rather good optical surfaces can strongly affect an interferometer.

- *Polarization effects.* The electric vector components respectively perpendicular (s) and parallel (p) to the plane of incidence to a flat mirror generally experience two different phase shifts at reflection. If an interferometer can be built so that the reflections in each arm follow the same sequence of changes in direction, and if the corresponding mirrors at each reflection are of the same type, then both beams will experience the same phase shifts, and the respective s and p components will combine independently in the focal plane and produce identical fringe packets. However if the sequence of reflections is different, s – p differences may occur. If the s – p shift between the two beams is $\phi_{\rm sp}$, the interferogram will have its visibility reduced by $|\cos(\phi_{\rm sp}/2)|$.

1.3.2 Observing through the turbulent atmosphere

Atmospheric turbulence is a major contributor to the difficulty of ground-based interferometry: atmospheric fluctuations cause visibility losses which can be more troublesome than instrumental losses because potentially they are larger in magnitude and variable in time.

Wavefront distortion

Inhomogeneities in the atmosphere's refractive index introduce rapidly varying random delays in the beam arrival times resulting in random motions of the fringe pattern with time. These are difficult to compensate as the atmospheric coherence time[6] is typically of order 5-20 msec in the optical (see Paresce [45]). In the infrared, these constraints are significantly relaxed, which explains why most of the new large interferometers have chosen this band.

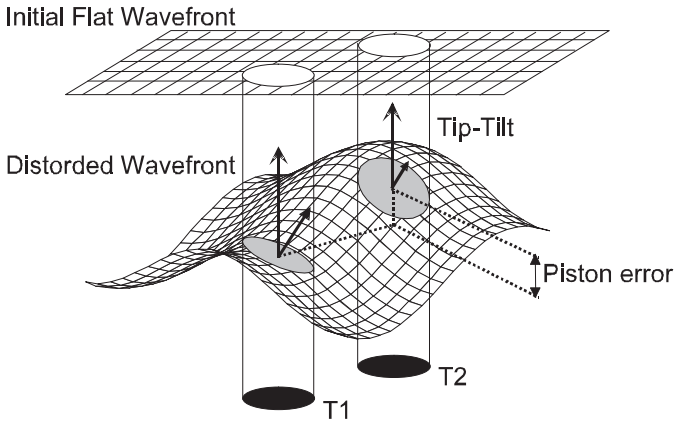


Figure 1.9: Wavefront distortion induced by atmospheric turbulence (from Paresce [45]).

The correction for this effect depends on the size of the individual telescopes and the wavelength band of operation. To understand this, consider the shape of the wavefront reaching the array as shown in figure 1.9. Above the atmosphere, the wavefront is perfectly flat as in the ideal case considered so far but, on the ground, the wavefront is distorted by the turbulent air that impresses on it corrugations with typical dimensions of a few centimeters in the optical up to several meters in the mid-IR. Thus, if the telescope apertures T1 and T2 are smaller than the size of the corrugations, they select small, relatively flat areas of the wavefront and one has only to correct for the slightly different orientations (tip/tilt) of the two wave portions with a fast steering mirror to obtain reasonably stable images that can be properly superimposed.

Correction of the wavefront

- *Adaptive optics.* For larger telescopes, the mirrors accept several wavefront corrugations simultaneously, producing a combined image containing many speckles on each of which is superimposed an interference pattern. Uncorrected, this residual wavefront distortion would severely limit the usefulness of large telescopes in an interferometric array and, therefore, the limiting sensitivity of these devices. Higher order adaptive optics compensation with deformable mirrors on each large telescope is the only way out of this dilemma.

[6]Characteristic time during which the phase shift induced by the atmospheric refraction index inhomogeneities may be considered as constant.

- *Fiber optics.* Single-mode fiber optics may be used within an interferometer (a) to essentially select the plane-wave part of a wavefront, (b) to split a guided wave into desired intensity ratio, and (c) to combine two guided waves. The advantage of single-mode fibers is that when only the plane-wave part of the wavefront is used, the fluctuations in visibility due to random warping of the wavefront (due to atmospheric fluctuations as well as optical surface corrugations) are dramatically reduced.

- *Fringe tracking.* Even with these corrections, the fringes would still move, however, due to the mean delay in the arrival times of the distorted waves (piston error shown in figure 1.9). This effect is usually partly compensated by using a fringe tracker that detects and follows the instantaneous position of the fringes in real time and feeds this information back to a fast optical compensator to "freeze" the fringe motion. This allows the device to integrate in time and significantly increases measurement precision. To work properly, however, the fringe tracker needs a relatively bright compact source, whose fringes can easily be detected above the instrumental noise.

- *Phase determination.* In principle, were one able to carry out all the corrections just described in a perfect manner, a pair of movable telescopes defining a single baseline of varying length and whose projection on the sky rotated in response to Earth's diurnal motion would be sufficient for imaging even a complex but relatively bright source by precisely measuring visibilities and phases for each baseline one after the other. Unfortunately, the atmospherically induced delay varies randomly with the location and spacing of the apertures preventing a determination of the true phase relations between the various sets of measurements. Two ways to overcome this problem have recently been devised. The first relies on a remarkable additive property (called "phase closure") of the interferometer output phases if three or more baselines are simultaneously used: the phase addition of the interferometric signals cancels the atmospherically induced phase shifts. In the second technique (called "phase-referencing"), the phase information from a reference object is used to determine the atmospheric phase, and correct the phase of the target source accordingly. The two objects must be sufficiently close together in the sky so that their beams traverse the same patch of atmosphere above the array.

Chapter 2

Detection of extra-solar planets

2.1 Context

In this section, a review of the techniques used to detect exoplanets is presented. Particular attention is given to Earth-like habitable planets, which are of main interest for ESA's IRSI-Darwin mission. In the following discussions, habitable (or "terrestrial") planets are defined as planets with a mass in the range $0.5 - 10\,M_\oplus$, leading to a radius in the interval $0.8 - 2.2\,R_\oplus$, and at surface temperature ($\sim$ 300 K) which allows the presence of liquid water.

Detecting and observing exoplanets is a rather tedious job, because of the two major difficulties explained below.

2.1.1 Dynamical range

The first difficulty is the huge contrast between the star and the planet. Indeed, in the visible and near-infrared range (0.4 to 3 μm), the planet almost only reflects the light emitted by the star, and the brightness ratio is of order 5×10^9 for an Earth-like planet around a Sun-like star. The contrast is more favorable in the mid-infrared, where the intrinsic thermal emission of the planet is maximal (see figure 2.1). In the following discussions, we will assume a temperature of $\sim$300 K for Earth-like exoplanets, giving a brightness ratio of about 10^7 between 7 and 20 μm. The magnitudes m and fluxes F_ν at 10 μm for a Sun-Earth system at 10 pc are of $m_p \simeq 20.7$ and $F_{\nu,p} \simeq 0.23$ μJy for the planet and of $m_* \simeq 3.6$ and $F_{\nu,*} \simeq 1.6$ Jy for the star (values from [18]).

2.1.2 Angular resolution

The second difficulty is that the angular separation between a planet and its star is typically of the order of a fraction of an arcsec. Earth-like exoplanets are supposed to be in the habitable zone surrounding the star, at about 1 AU from a G-type star[1]. For example,

[1]The preliminary target list for Darwin ranges from F-type stars (habitable zone between 1 and 3 AU for a $1.5M_\odot$ star) to M-type stars (habitable zone between 0.1 and 0.3 AU for a $0.3M_\odot$ star). The greenhouse effect of CO_2 clouds could push the outer boundary a little farther.

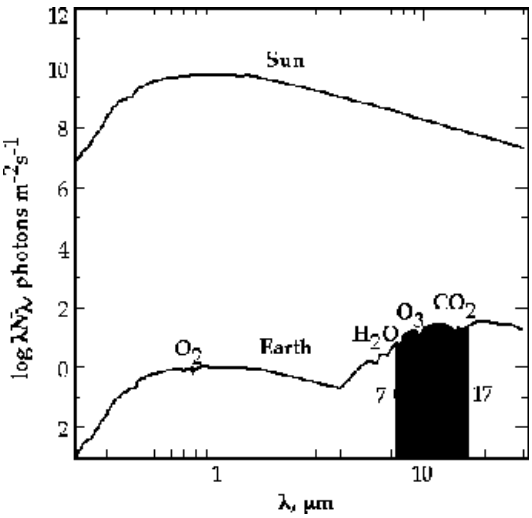


Figure 2.1: Fluxes of the Earth and the Sun as seen from 4 pc, from Angel *et al.* [2].

a terrestrial planet orbiting a solar-type star at 10 pc from the observer is seen with an angular separation of 100 mas. High angular resolution can be achieved with interferometry, as demonstrated in the first chapter: a 100 mas angular resolution is reached with a 20 m baseline at 10 μm. On the other hand a single aperture of about 1 m in diameter would be sufficient in the visible range to achieve such a resolution.

2.2 Indirect detection

The two difficulties mentioned above have forced the astronomers to search for exoplanets in a different way than direct imaging. Indirect methods are based on the influence of the planet on the observables of the stars. Thus, in this section, we only observe the star, and not the planet. It is the reason why those methods are called "indirect".

In the following discussions, we will adopt the notations of Perryman [46], who gives an excellent review of extra-solar planet detection. Parameters used are mass M, radius R, and luminosity L, with subscripts $*$ and p referring to star and planet respectively. Systems are characterized by their orbital period P, semi-major axis a of the planet's orbit, eccentricity e, inclination i of the orbital plane with respect to the plane of the sky ($i = 0°$ face-on, $i = 90°$ edge-on), and distance d from our Solar System. Unless explicitly noted, we will assume a single planet orbiting the star.

2.2.1 Radial velocity

This method is the one that first allowed secure detection of planets, by Mayor and Queloz [36], around 51 Peg. Since 1995, more than 60 planets have been detected with radial velocity measurements.

Principle

The gravitational pull from a planet on its host star can be detected by precise measurements of the radial velocity of the star by means of spectroscopy, since the star wobble induces a periodic Doppler shift on the stellar spectrum. The third Kepler law

$$a^3 = \frac{G}{4\pi^2} P^2 (M_* + M_p) \simeq \frac{G}{4\pi^2} P^2 M_* \,, \tag{2.1}$$

where G is the gravitational constant, must be restated in our case because we only observe the star, orbiting around the system's barycenter (Kepler used to watch the planets, orbiting a fixed star). Replacing $G(M_* + M_p)$ by the so-called "mass function" $\mu = GM_p^3/(M_* + M_p)^2$ gives a relation between the masses, the orbital period and the maximum radial velocity[2] $v_{max} = 2\pi a \sin i/P$:

$$\frac{(M_p \sin i)^3}{(M_* + M_p)^2} = \frac{v_{max}^3 P}{2\pi G} \,, \tag{2.2}$$

which is only valid for small eccentricities. For example, measurements of v_{max} (59 m/s) and P (4.23 days) for 51 Peg (figure 2.2) lead to $M_p \sin i \simeq (1.36 \times 10^{20} M_*^2)^{1/3}$, with a good approximation since $e = 0.01$ (quasi-circular orbit). As the $\sin i$ factor is a priori unknown, the mass of the planet remains undetermined. In fact, the radial velocity technique only allows a minimum mass determination, obtained for $\sin i = 1$ (minimum mass of 0.5 M_J for 51 Peg's companion). Note that orbital systems seen face on ($\sin i = 0$) result in no measurable radial velocity perturbation.

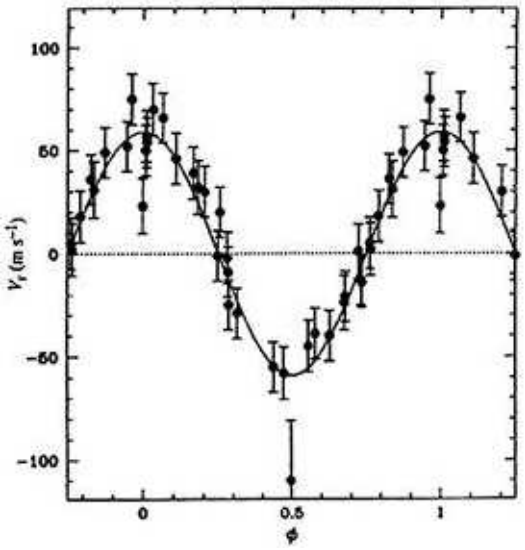


Figure 2.2: Radial velocity versus orbital phase for 51 Peg (Mayor and Queloz [36]).

[2]Writing so, we assume a quasi-circular orbit, which seems to be a quite frequent case.

In the more general case of a planetary orbit with non-zero eccentricity, the velocity amplitude v_{max} is given by

$$v_{max} = \left(\frac{2\pi G}{P}\right)^{1/3} \frac{M_p \sin i}{(M_* + M_p)^{2/3}} \frac{1}{(1-e)^{1/2}} \ , \tag{2.3}$$

which reduces to (2.2) for $e = 0$. This equation shows that radial velocity measurements favor the detection of systems with massive planets and small periods P (and hence small semi-major axes a). Velocity variations do not have a nice sinusoidal shape in this more general case.

Limitations

Detectability depends on the number of orbital cycles observed and on the number of observations per cycle. If many cycles are observed (possible when P is small), Fourier techniques allow measurements of velocities comparable to the Doppler errors. On the other hand, if the planetary period is much larger than the observing time, secure detection is only possible for amplitudes at least 4 times larger than the errors. The current precision reaches about 3 m/s, and a projected precision of 1 m/s is under development, for example with the HARPS instrument at La Silla (Chile), which will achieve a long-term precision of 1 m/s for differential radial velocity (not before fall 2002).

Moreover, the Doppler technique is intrinsically limited to around 1 m/s because of the effect of star spots and convective inhomogeneities, which create a proper velocity field on the stellar surface. In contrast, the motion of the Sun caused by the Earth has a semi-amplitude of 0.1 m/s, and therefore detection of Earth-like exoplanets is definitely not achievable with radial velocities. The minimum detectable mass is about 10 $M_\oplus / \sin i$ for a planet at 1 AU from a one solar-mass star.

Other limitations are signal-to-noise considerations, which limit observations to the brightest stars (typically $V < 8$ at the moment), and also the $\sin i$ factor, which could be estimated with other techniques, such as photometric transits, or constrained by other physical arguments.

2.2.2 Astrometric position

Principle

The gravitational signature of a planet on its host star can also be detected perpendicularly to the line-of-sight by using astrometric measurements. The path of a star orbiting a star-planet barycenter appears as an ellipse with semi-major axis α given by:

$$\alpha = \frac{M_p}{M_*} \cdot \frac{a}{d} \tag{2.4}$$

where α is in arcsec when the planetary semi-major axis a is in AU and d in parsec (Perryman [46]). Precise astrometric measurements of α, achieved with differential methods on ground based interferometers, and combined with an estimation of a by means of the

period, directly give us the mass of the planet without the $\sin i$ undetermination, if the star's mass can be estimated from its spectral type and d from the star's parallax motion. Astrometric measurements can also provide orbital constraints on $\sin i$. For multi-planet systems, the relative orbital inclination can be determined.

Seen from a distance of 10 pc, the Sun would oscillate with a typical relative displacement of 1 mas, principally due to Jupiter (figure 2.3). The Earth-induced displacement would only be 0.3 μas. The astrometric method privileges high masses, as radial velocity, but is more sensitive to large semi-major axes (i.e., long orbital periods).

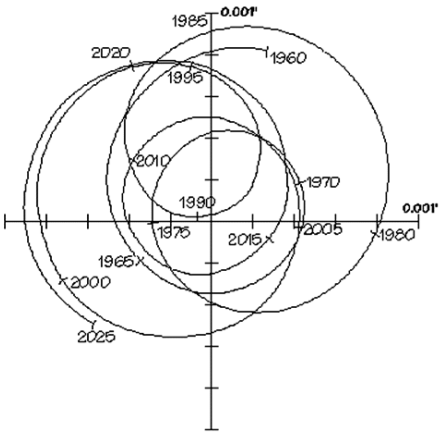


Figure 2.3: The wobble of the Sun's projected position due to the influence of all planets in the Solar System as it would appear from 10 pc (axes labeled in arcsec). Taken from the ExNPS Road Map web-site.

A 100 μas precision is currently achieved at radio and optical wavelengths, respectively with very long baseline interferometry (VLBI) and with the Palomar Testbed Interferometer, but no planet detections have been confirmed yet using this method[3]. The VLT and Keck interferometers are designed to achieve 10 μas precision. This is not enough for habitable planet search.

Limitations

Due to phase fluctuations of Earth's atmosphere, ground based astrometric measurements are limited to a maximal precision of about 10 μas. This high precision is only possible with very small fields of view (typically some tens of arcsec), where the atmosphere becomes essentially common mode[4]. If one wishes to make astrometric measurements with an accuracy significantly better than 10 μas (necessary for Earth-like planet detection), the

[3]From the Allegheny Observatory astrometric work done by George Gatewood, it appears that there are possibly two "classical jovians" orbiting Lalande 21185. These planets will hopefully hold up as the first confirmed astrometric detection of a planetary system.

[4]Atmospheric turbulence causes the position of a star to fluctuate with time. But if two stars are sufficiently close together in the sky, the motion of the two stars will be almost identical because the light from the two stars traverse almost identical paths through the atmosphere.

only solution is to go to space. For the planned Space Interferometry Mission (SIM), which will measure angles as small as 2 μas, a minimum planet of mass of 6.6 $M_\oplus$ could be detected in a 1 year orbit around a 1 $M_\odot$ star that is 10 pc from the Earth.

The ultimate limit to Earth-like planet detection from space is once again related to the non-uniformity of illumination over the disk of the star. Some sunspots have areas up to 1% of the area of the Sun and so can cause the apparent center of light to move by as much as 0.5% of the stellar diameter, whereas the Earth only induces a wobble with an amplitude of 0.03% of the stellar diameter (Woolf and Angel [60]). Young stars are even more active, with spots as large as 10% of the stellar disk. Detection of Earth-like planets is therefore not achievable with astrometric measurements (unless the star is very quiet).

Generalization

In the case of "hot Jupiters", the planet light can be "indirectly" detected by measuring the star-planet photocenter displacement when changing the observation wavelength (the planet being more luminous in the mid-IR range). The Keck interferometer, for instance, will perform two-color phase reference interferometry with the capability to detect Jupiter sized planets at a close separation of 0.15 AU to parent stars at a distance of 10 pc.

2.2.3 Photometric transits

Principle

For planetary systems oriented by chance with their orbital plane including the line-of-sight to the observer, occultations will reveal the presence of planets. The probability to detect a transit depends on the orbital system orientation, the orbital distance and the stellar diameter. To detect an occultation, the orbital inclination i must satisfy the following relation:

$$\cos i < \frac{R_*}{a} \, , \tag{2.5}$$

which gives a probability $p \simeq R_*/a$ to observe the transit for a randomly oriented system ($p \simeq 0.005$ for a Sun-Earth system). Under the assumption of uniform surface brightness for the star[5], the luminosity drop ΔL produced by an object along the line-of-sight is given by the surface ratio

$$\frac{\Delta L}{L_*} = \left(\frac{R_p}{R_*} \right)^2 \, , \tag{2.6}$$

and, if δ is the latitude of the transit on the stellar disk, the duration of the transit reads

$$D_T = \frac{P}{\pi} \left(\frac{R_* \cos \delta + R_p}{a} \right) \, . \tag{2.7}$$

[5]Strictly speaking, the effect includes a dependence on the local surface brightness of the stellar disk, which varies with radius due to limb darkening.

For Jupiter-like and Earth-like planets, the relative luminosity drop is respectively 10^{-2} and 8.4×10^{-5}, while the transit duration equals 25 hours and 13 hours. The highest sensitivity will be for larger planets in closer orbits (as for radial velocity).

Photometric transit is a very simple method (it only requires a precise photometer), which provides very valuable information on the orbital system: from the luminosity drop and an estimation of the star's radius (e.g. from spectral classification), we can infer the radius of the planet. From the periodicity of the transits, we can deduce the semi-major axis by means of Kepler's third law. Moreover, from the duration of the transit, we can estimate the latitude δ, and hence the orbital inclination from $\cos i = (R_* \sin \delta)/a$.

Recently, a first event has been detected by Charbonneau *et al.* [17], who used the complementarity between photometric transits and radial velocity. As shown above, the radial velocity technique is more sensitive to large $\sin i$, and is susceptible to detect edge-on planetary systems. Charbonneau has observed two transits on star HD 209458, in perfect agreement with radial velocity observations (figure 2.4).

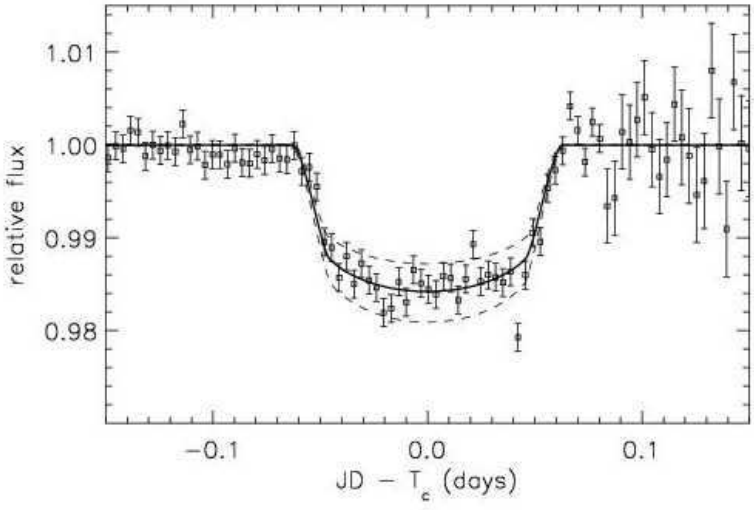


Figure 2.4: Photometric variations of HD 209458 due to a planetary transit (from [17]).

Limitations

The principal disadvantage of the method is the very weak probability of such events (the inclination i must be very close to 90°). To overcome this limitation, we can make a pre-selection of target stars whose planetary companion is likely to have its orbital plane perpendicular to the plane of the sky. Another solution is to perform photometric surveys on large populations of stars, as will be made with COROT or Eddington.

Ground-based photometry is limited to a 0.1% precision due to atmospheric extinction, and to about 0.01% by scintillation. So, ground-based detection of Earth-like planets could only be possible on the very best astronomical sites. Even in those sites, long uninterrupted observations ($\gtrsim$ 13 h) are not possible. Dedicated spacecraft missions are needed to avoid those limitations. COROT (planned for launch in 2004), and later Eddington, will perform

photometric surveys on a large population of stars with high accuracy over large fields of view (typically 6 square degrees). A photometric precision of 10^{-6} will be reached, allowing detection of Earth-like extra-solar planets.

Just as for the previous techniques, the fundamental limitation comes from stellar variability. Intrinsic stellar variations, flares, sunspots, etc. could cause photometric variations of the same order as planetary transits in the case of Earth-like planets. Besides those limitations, the method is presently considered as one of the most promising means of detecting planets with masses significantly smaller than that of Jupiter.

Generalization

For giant planets very close to their host star, an occultation is not needed to observe luminosity variations. The reflected light intensity $L_p \propto a^{-2} R_p^2 L_*$ can indeed create a detectable modulation of the total received light, even if the planet cannot be imaged as such. This may also occur if the orbital plane is somewhat inclined to the line-of-sight (see Perryman [46]). Such measurements would enable an evaluation of the planet albedo.

2.2.4 Gravitational microlensing

Principle

Gravitational lensing is the focusing and hence amplification of light rays from a distant source by an intervening object. This effect has first been predicted by Einstein in 1912. The relative motion between the background source, the intervening lens and the observer will lead to an apparent brightening and subsequent dimming of the resulting image, which may occur over time scales of hours and upwards. One speaks of "microlensing" effect when the multiplicity of the subsequent image is not resolved. This happens when the deflector's mass is too small, so that the angular deviation of the light is much smaller than the angular size of the telescope's diffraction pattern.

Events are characterized in terms of the Einstein ring radius

$$R_E = \left(\frac{4GM_L}{c^2} \frac{(D_S - D_L)D_L}{D_S} \right)^{1/2} , \tag{2.8}$$

where M_L is the mass of the lensing object, D_L and D_S the distances to the lens and to the source. The Einstein angle is then defined as $\theta_E = R_E/D_S$, and microlensing magnification as a function of time reads (see Perryman [46])

$$A(t) = \frac{u^2(t) + 2}{u(t)\sqrt{u^2(t) + 4}} , \tag{2.9}$$

where $u(t) = \theta_S/\theta_E$ is the angular distance between the deflector and the source in units of the Einstein angle. For a single lensing star, the light curve is symmetrical, and centered on the star position. An infinite amplification only arises when the lens and the source are perfectly aligned, and when the source is infinitesimal. The caustic for a single lens is the single point behind the lens, and the subsequent image is the Einstein ring.

Now, if the lens consists in two point-like objects (e.g. a star and its orbiting planet), the caustic is not point-like any more, and depends on the mass ratio M_p/M_* and on the projected planet-star distance. With the addition of a planet around the central star, the light curve of the background source will be very close to that of the single lens for most of its duration, but with fine structure comprising additional sharp peaks (see figure 2.5).

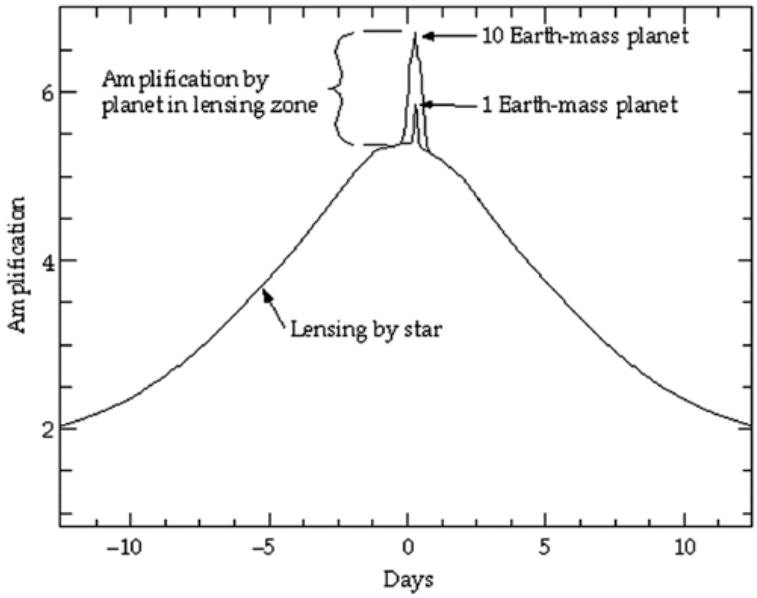


Figure 2.5: Theoretical amplification curve for a microlensing event in the case of a double lens (star and planet). Two different planet masses are represented. The "lensing zone" is the distance range in which the planet produces caustics inside the Einstein ring of the star. This zone forfuitously corresponds to star/planet distances of a few AU.

Massive observational programs capable of surveying millions of stars have been underway since 1993, and have provided the first microlensing event detections. Scientific teams such as MPS (Microlensing Planet Search) and EXPORT (Extrasolar Planet Observational Research Team) gather detailed photometric information about lensing events sifted from the vast survey data streams. The event MACHO 97-BMG-41 was recently reported as the first convincing example of planetary microlensing (Bennett *et al.* 1999, [10]). A triple lens model, where the lens consists of a planet orbiting a binary star, can be fitted to the observation data (see figure 2.6).

Limitations

Unlike other methods, microlensing is not really limited by the target system distance since it requires no photon from either the planet or the parent star. Moreover, microlensing is sensitive to low mass planetary system, and is the only known ground-based method to probe Earth-mass planets orbiting around main sequence stars.

But on the other hand, precise alignment is required for a detectable brightening. The chance of substantial microlensing magnification is extremely small, $\sim 10^{-6}$ for background

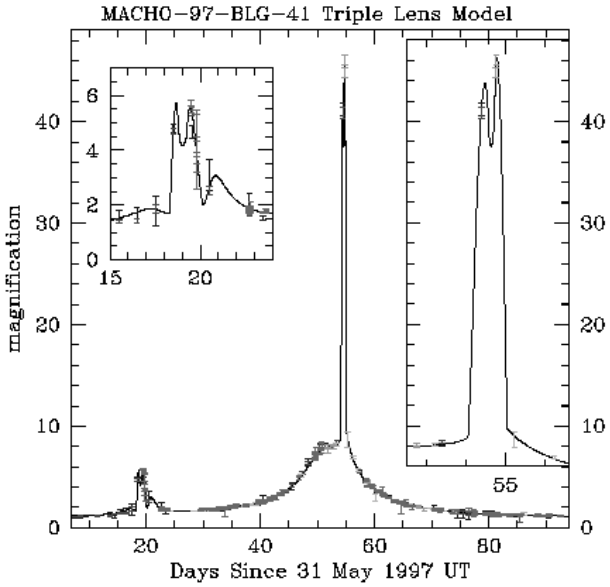


Figure 2.6: The best fit light curve of the MACHO-97-BLG-41 microlensing event, from Bennett *et al.* [10]. Note that circumbinary planetary systems are a unique territory of the microlensing planet search technique.

stars in the Galactic bulge, nearby Magellanic Clouds, or the nearby spiral galaxy M31. Only massive observational programs can provide the required data for planet search. Another drawback is that specific systems cannot be selected for study, and once a microlensing event is detected, an independent event is unlikely to recur for the same system on any relevant time scale. Moreover, measurements only give the planet/star mass ratio and their projected separation (in units of the Einstein radius of the lens).

2.2.5 Pulsar timing

Rapidly spinning highly-magnetized neutron stars are formed during the core collapse of massive stars (8-20 $M_{\odot}$) in a supernova explosion. These neutron stars, called pulsars, emit narrow beams of radio emission parallel to their magnetic dipole axis, seen as intense pulses at the object's spin frequency. This frequency ranges from about a second for "normal" pulsars to about a millisecond for old pulsars (principally found in binary systems).

If a pulsar with period T_* has a planetary companion, the gravitational influence of the planet induces a wobble on the pulsar, and hence a variation of the observed pulsar period since the light travel distance is continuously changing. This variation is expressed as a function of the orbital plane inclination i and of the pulsar orbit semi-major axis a_*:

$$\delta T_* = \frac{a_* \sin i}{c} \,, \tag{2.10}$$

where c stands for the speed of light. The high accuracy of pulsar timing (a few microseconds) and the high stability of pulsar rotation allow detection of lower mass bodies orbiting

the pulsar. Jovian or terrestrial planets are expected to be detectable around normal slow pulsars, while substantially lower masses, down to that of our Moon and largest asteroids, could be recognized in millisecond pulsar timing residuals (Perryman [46]).

Detection of several Earth-mass planets have been reported since 1992, when a first planetary system has been discovered around an object other than our Sun (Wolszczan [57]). However the environment of old millisecond pulsars is not particularly well suited for life growth, due to strong electro-magnetic radiation. This is the reason why pulsar timing does not give rise to as much interest as other detection techniques. But it can still provide constraints on plausible formation processes (see Perryman [46] for more details).

2.3 Direct detection

2.3.1 Basic considerations

Direct detection methods, in which we distinguish radiation that comes from the planet, seem to offer the best way to discover Earth-like planets. Such methods are also the only way to determine physical characteristics such as temperature and chemical composition through spectroscopy. It is therefore the only way to assess the habitable character of Earth-like planets. Note that a clear distinction exists between the goal of direct detection, which simply requires resolving the planet's radiation from that of the star, and the goal of resolving the disk and mapping the surface of an Earth-like planet.

Single classical telescope

Let us first consider the possibility to directly detect extra-solar planets with a ground-based or space-borne telescope observing either in the visible or in the mid-infrared range. Our first priority is to devise an optical system that will resolve objects at a typical angular separation $\theta_p = 0.1$ arcsec[6], cleanly enough to allow planet detection. Appropriate instrument scale sizes, λ/θ_p, range from about 2 m in the optical and near-infrared to about 20 m at 10 μm. These telescope sizes are conceivable nowadays, even if an interferometer would probably be preferred in the mid-infrared range.

On the other hand, the criterion for resolving and detecting objects of enormously different intensities needs some discussion. The planet does not have to be brighter than the local halo of scattered starlight, but its detection does require that, at a minimum, the random fluctuations in the halo due to photon noise be smaller than the planet signal. Defining the gain of an optical system as the ratio of the stellar peak intensity to the star intensity at the image of the planet, Woolf and Angel [60] derive typical values of the gain for a Sun-Earth system at 10 pc: about 6×10^8 at 800 nm (e.g. for O_2 spectroscopy with resolution $R \sim 40$), but only about 9×10^3 at 9.7 μm (e.g. for O_3 spectroscopy with resolution $R \sim 10$). These values were computed for a signal-to-noise ratio (SNR) of 5 in 10 hours for detection, or equivalently, for a SNR of 25 in 10 days for spectroscopy.

[6]This angular separation corresponds to the case of a Sun-Earth system seen at 10 pc.

Unfortunately, the point-spread function for usual circular telescope aperture drops only slowly with angle. The gain at an angle θ far from the central peak is given by

$$G(\theta) = 4\pi^2(D\theta/\lambda)^2 \,, \tag{2.11}$$

where D is the telescope diameter. For example, for $\theta = 10\lambda/D$, the gain is about 5×10^4. For the G values derived above, we find that the diameter required for infrared detection ($D = 10\lambda/\theta_p \sim 200$ m) is too large to be practical, while for any realistic diameter, the gain is far too low for optical planet detection.

Adaptive optics on large ground-based telescopes

When ground-based observation is considered, the planet signal is immersed in the seeing profile arising from turbulent atmospheric refraction, which is far more problematic than the diffraction profile. Prospects for ground-based planetary imaging have concentrated on the use of adaptive optics (see Angel [1]). This method relies on continuous measurement of the wavefront from a reference star (or from an artificial laser guide star), then applying an equal but opposite correction using a deformable mirror containing actuators distributed across its surface, at frequencies of order 1 kHz.

Adaptive optics programs underway at the world's largest ground-based telescopes (VLT, Keck, LBT) may ultimately have the sensitivity to detect giant planets around some tens of nearby stars. In his 1994 article, Angel describes this possibility in a very detailed way. His conclusion is that Jupiter-like planets with a 5-AU orbit around nearby stars could be detected with 8-m class telescopes in a 1-hour integration at the 5σ level, if the actuator spacing was reduced below 5 cm. More than 10^4 actuators would be needed, with a sampling frequency of about 2 kHz. This optical system would even be capable of detecting an Earth-like planet around the singularly close binary star α Centauri (1.3 pc). But a more comprehensive search for Earth-like planets would require massive optical aperture, such as the 100-m OWL telescope (under consideration), equipped with about 10^6 actuators (see Perryman [46]).

Apodization

Apodization uses controlled variations in transmission or reflectivity of the entrance pupil in order to change the side-lobe intensities of the diffraction pattern. To obtain a strong suppression close to the star image, a new form of apodization has been devised by Angel *et al.* [2]. An annular pupil mask has the effect of deepening and broadening the first dark ring of the diffraction pattern. Angel *et al.* consider the possibility of imaging the thermal emission of the planet in order to reduce requirements for gain and hence surface accuracy. Gain greater than 10^5 is achievable over the range of radii $\theta = 1.7$ to 2.3 λ/D. A space telescope designed to sense a 1-AU planet at 5 pc at 10 μm wavelength would require a diameter of 20 m, which is still uncomfortably large.

Conclusion

In the visible range, the main limitation to imaging devices is the huge contrast, which cannot be fought with "classical" telescopes. On the other hand, when observing in the

mid-IR range, the main limitation comes from the angular resolution, which forces us to use very large apertures. Therefore there are two kinds of possibilities for direct exoplanet detection, depending on the spectral range. The first consists in optical single-aperture coronagraphy, while the second relies on infrared interferometry.

2.3.2 Classical Lyot coronagraphy

Coronagraphy is a rather old technique used to observe highly contrasted objects. Lyot has originally devised it at the beginning of the 1930's in order to observe the Solar corona without eclipse. He managed to reduce the scattered light to a maximum by re-imaging the primary focal plane and by placing a so-called Lyot mask in the associated pupil plane (see figure 2.7). This technique allows to reduce the solar halo by a factor between 10 and 100 when well-polished optics are used (see Ollivier [43]). Super-smooth optics are essential in order to reduce the light scattered by the surface roughness of optical elements.

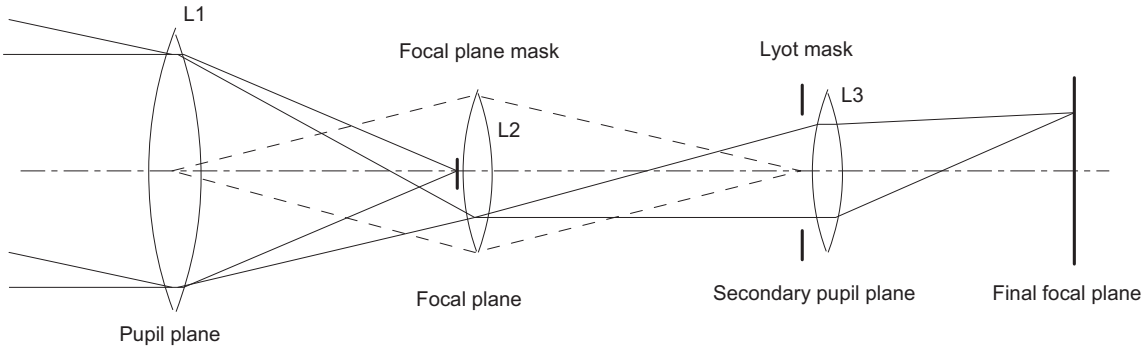


Figure 2.7: Principle of a Lyot coronagraph.

This technique has been adapted to stellar coronagraphy with great success, particularly when coupled to adaptive optics systems (in order to reduce the light scattered by the atmosphere) or when used on space telescopes. Unfortunately, classical Lyot coronagraphy does not allow very close-sensing of the star environment: the smaller the focal mask, the wider the scattering blur, and the more one has to stop the pupil down to eliminate light diffracted by the edge. In practice, the central area of the field is masked up to more than 3 Airy rings, providing sensing capabilities not better than 100 mas for a 8-m class telescope in the visible range. So, even the world's largest telescopes do not allow the study of the habitable zone for typical G-type stars at 10 pc with a Lyot coronagraph. Moreover, by hiding the central part of the field, we do not benefit from the region where adaptive optics is the most efficient. New types of stellar coronagraphs have recently been proposed to avoid this problem.

2.3.3 Phase mask coronagraphy

Principle

This coronagraph, proposed by Roddier and Roddier [49] is essentially identical to the Lyot coronagraph apart from the stellar occulting mask which is replaced with a phase mask. The phase mask is much smaller than the occulting mask: it covers only a fraction of the core of the Airy pattern. It is totally transparent, but induces a 180° phase shift on the incoming wave. After transmission through the mask, the complex amplitude can be expressed as a sum of two terms[7] $\mathbf{A}(\vec{\theta}) + \mathbf{B}(\vec{\theta})$, where $\mathbf{A}(\vec{\theta})$ is the complex amplitude of the Airy pattern in which the central part has been removed, and $\mathbf{B}(\vec{\theta})$ is a function equal to zero everywhere except over the small mask where it is negative (180° phase shift), see figure 2.8.

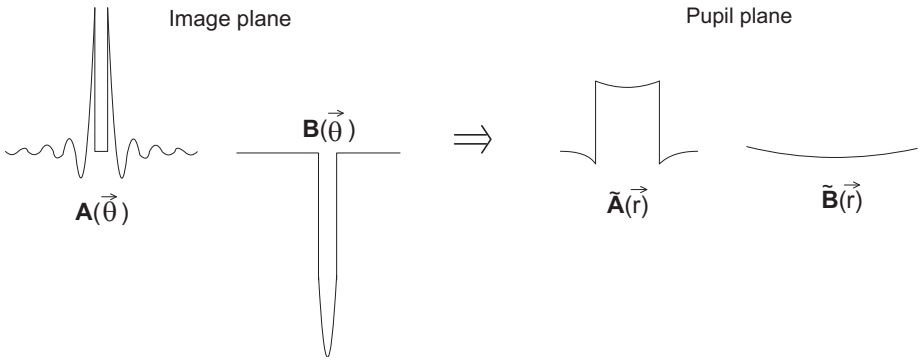


Figure 2.8: Sketches of the amplitude in the image plane (left) and in the following pupil plane (right), adapted from Roddier [49].

The complex amplitude in the following pupil plane is the Fourier transform $\widetilde{\mathbf{A}}(\vec{r})+\widetilde{\mathbf{B}}(\vec{r})$ of the previous sum, where the position $\vec{r}$ is expressed in wavelength unit in the telescope aperture plane. Since the Airy pattern has been set to zero at the origin, its Fourier transform $\widetilde{\mathbf{A}}(\vec{r})$ resembles the telescope aperture transmission function but it is zero mean, i.e., $\widetilde{\mathbf{A}}(\vec{r})$ is positive inside the telescope aperture, but negative outside. If $\mathbf{B}(\vec{\theta})$ is much narrower than the Airy pattern, its Fourier transform $\widetilde{\mathbf{B}}(\vec{r})$ is much wider than the telescope aperture and negative everywhere (see figure 2.8). If its modulus can be made equal to that of $\widetilde{\mathbf{A}}(\vec{r})$ inside the aperture, both terms will cancel out by destructive interference sending all the star light outside the telescope aperture where it is intercepted by the Lyot stop. On the other hand, light from a faint companion outside the phase mask will be little affected and almost totally transmitted.

There are two possibilities to balance the two illuminations: if a very small phase mask is used, the light outside the phase mask can be reduced by a density filter[8], but the

[7] The position vector $\vec{\theta}$ is considered here as angular coordinates on the sky.

[8] Theoretically, light rejection can be as good as one wants by making the mask smaller and hence the two amplitudes more uniform over the telescope aperture, but at the cost of an ever increasing exposure time.

phase mask can also be extended in order to cover exactly 50% of the energy in the Airy pattern. In that case, the two interfering beams are no longer uniform over the aperture plane, but computer simulations show that the rejection level is still better than that of a conventional Lyot coronagraph. In addition, phase mask coronagraphy allows observations to be made much closer to the central star, and therefore detection of close companions. A drawback of the technique is that the phase mask is wavelength dependent, as well as the width of the Airy disk. This restricts the observation to small spectral bins (typically $\Delta\lambda \sim 0.03$ μm for λ about 1 μm). Means to produce achromatic phase masks are currently under development and could solve the first problem in a near future (see section 3.4.4). Glass materials with an appropriate color dependent transmission could be used to balance the fluxes over a wide spectral range. Another suggestion is to use a Bragg hologram as a phase mask: the mask could be made so that each wavelength sees it as being the right size and having the right phase shift.

Performances

Using a 3.6 m telescope and an observation wavelength of 1.65 μm, Roddier [49] shows that the attenuation of the envelope at 0.4 arcsec can be as high as a factor of 210 for a phase mask covering 43% of the Airy pattern[9] (i.e., 50% of its energy), and as high as 10^3 for a small phase mask, with radius 0.2 times that of the Airy ring (illumination is reduced by a factor about 38 by means of a density filter in that case). With these performances, brown dwarf companions can easily be detected very close to their host star, but this is clearly not enough in order to detect Earth-like exoplanets. Moreover, performances of ground-based coronagraphs strongly depend on the quality of the adaptive optics system, because it determines the amount of atmosphere-scattered light.

Improvements

The first possible improvement comes from apodization: if the entrance pupil has a negative curvature (transmission decreases towards the center of the pupil), the positive curvature of $\widetilde{\mathbf{A}}(\vec{r}) + \widetilde{\mathbf{B}}(\vec{r})$ can be cancelled. Guyon and Roddier [20] have developed an algorithm that computes the optimal transmission map for the pupil. This optimal map has a 49% minimal transmission at the edges of the pupil and a total integrated transmission of 73%. The theoretical extinction factor increases from 210 to 10^7 for a monochromatic on-axis point source.

The ultimate limitations of space-borne coronagraphic systems do not come from their optical performances, but from scattering due to imperfections in the optical surfaces of the collecting system. In order to reduce these imperfections, Malbet *et al.* [35] have proposed the use of a deformable mirror to decrease the scattering level in local regions called "dark holes". Since scattering close to the optical axis comes from the figure errors of the mirror with large spatial scale, adaptive optics for space telescopes could provide a method of decreasing the on-axis scattering level. This demands the deformable mirror to be densely-actuated. Another way to improve the performances of coronagraphic systems is the use

[9] This rejection ratio is reduced to 156 when one takes into account the loss through the Lyot stop.

of dark-speckle analysis (see section 2.3.5).

The adaptation of a phase-mask coronagraph to the NGST is now being considered (see Boccaletti *et al.* [12]). The instrument concept relies on the combination of phase-mask coronagraphy, adaptive optics (dark-hole method) and dark-speckle analysis as the last stage of starlight suppression. Its ultimate goal is to attain 10^9 dynamic range in a 1 arcsec field of view for the detection of Earth-like planets around nearby stars in the visible and near-IR ranges. This dynamic range would be achieved in a few tens of hours.

Four-quadrant phase-mask coronagraphy

A new type of phase-mask coronagraph has recently been proposed by Rouan *et al.* [50]. It uses a four-quadrant binary phase mask (0, π) covering the full field of view in the focal plane. The mutually destructive interferences of coherent light from the main source produce a very efficient nulling. The computed rejection rate of this coronagraph appears to be very high since, when perfectly aligned and phase-error free, it could in principle reduce the total amount of light from the bright source by a factor of 10^8 at the location of the first Airy ring relative to the Airy peak. When used on Earth, this coronagraph appears to be less sensitive to atmospheric turbulence and has a larger dynamic range than the traditional phase-mask coronagraph. Simulations have shown that a ground-based detection at a contrast of 10 magnitudes between stars and faint companions is achievable under excellent atmospheric conditions.

2.3.4 Achromatic interfero coronagraphy

Principle

The basic principle of an achromatic interfero coronagraph relies on a Michelson interferometer modified by inserting in one arm an achromatic π-phase shift and a pupil π-rotation in its plane (Gay and Rabbia [19], Baudoz *et al.* [5]). Both effects are obtained by replacing in one arm the flat mirror by a cat's eye (see figure 2.9). The two beams are recombined at a common pupil plane where a focusing element forms an image on a camera. When the source is point-like and set on-axis, the coronagraph delivers a point spread function (PSF) whose maximal intensity is weighted at zero when the optical path difference (OPD) is set to zero, so that the image plane is utterly dark. This nulling would also affect the planet, but because of pupil rotation the light from the companion is unaffected by the destructive interference process. In that case, the off-axis source yields two twin images (half power each) in the focal plane while the "blinding" central on-axis source has been removed thanks to the nulling process.

Mathematically, if $P(\vec{r})$ is the entrance pupil function and $\phi(\vec{r})$ the phase of the incoming wave (with $\vec{r}$ the coordinate in the pupil plane), the output amplitude distribution writes

$$\mathbf{A}(\vec{r}) \propto rt \left(P(\vec{r}) e^{i\phi(\vec{r})} - P(-\vec{r}) e^{-i\phi(\vec{r})+i\pi} \right) \tag{2.12}$$

where r and t are the reflection and transmission coefficients of the beam-splitter. Assuming perfect symmetry for the entrance pupil ($P(\vec{r}) = P(-\vec{r})$), this expression reduces to

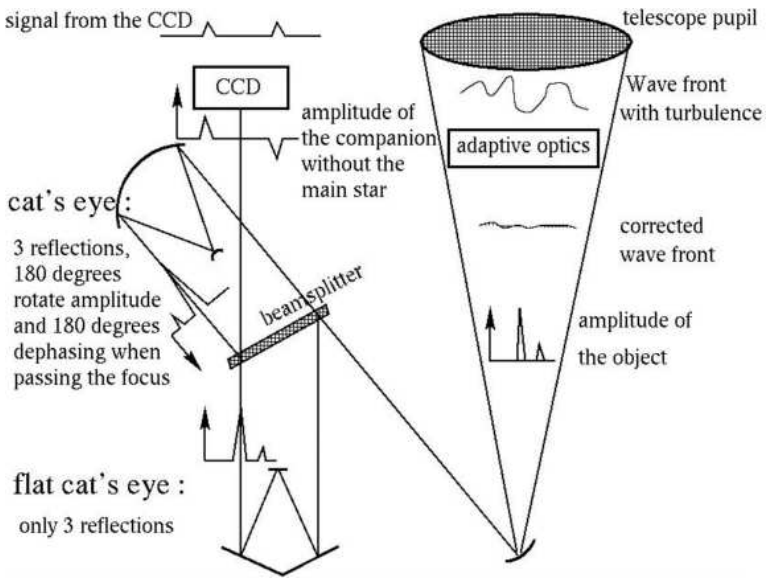


Figure 2.9: Principle of the AIC (figure from Pierre Baudoz' personal web-site). The corrected wavefront is divided in two equal parts by the beam-splitter. One of the beams undergoes a 180° rotation and phase shift by passing through a focus. The resulting image consists in two symmetrical replica of the off-axis planet.

$2irtP(\vec{r})\sin(2\pi\vec{r}\cdot\vec{\theta_0})$ since the phase of the incoming wave from a point source located at $\vec{\theta_0}$ simply reads $\phi(\vec{r}) = 2\pi\vec{r}\cdot\vec{\theta_0}$. To this pupil plane distribution corresponds the following amplitude distribution in the image plane (obtained by a Fourier transform):

$$\widetilde{\mathbf{A}}(\vec{\theta_f}) \propto 2irt(\widetilde{P}(\vec{\theta_f} - \vec{\theta_0}) - \widetilde{P}(\vec{\theta_f} + \vec{\theta_0}))\ , \tag{2.13}$$

which cancels everywhere on the image plane if the source is centered on $\vec{\theta_0} = 0$. On the other hand, any off-axis source produces two symmetrical images.

Because of its original design, this coronagraph has some specific properties. First, the achromaticity of the phase shift results from the crossing of an additional focus. A second characteristic is that the interferometric process does not result in a "dark zone" at the center of the image, but rather suppresses in the whole image plane all light coming from an on-axis unresolved source. Two crucial features are the need for perfect pupil symmetry, which could be broken by residual aberrations of the telescope (or even by the spider) or by the atmospheric turbulence, and the need for an accurate OPD control, which can be achieved by measuring the constructive output of the beam-splitter (the non-science output where the on-axis light is sent). Another specific feature of the AIC relates to close-sensing capabilities since it allows sensing as close as a fraction (0.5 to 0.3) of the Airy radius, which is better than the diffraction limit set by the aperture and not reached by other coronagraphs.

Performances

As in the case of phase-mask coronagraphy, the performances of a ground-based AIC depends on the quality of the adaptive optics system. The first observations with an AIC took place in October 1997 at the 1.52 m telescope of the Observatoire de Haute Provence, equipped with the BOA adaptive optics system. Working in the K band ($\lambda = 2.2$ μm), a magnitude difference ΔK of about 4.5 is detectable at one Airy radius from axis, with the help of mean radial profile subtraction [7]. On the other hand, theoretical simulations show that a magnitude difference about 12 can be reached in the case of complete adaptive optics correction up to the ninth radial mode.

Observing from space, the ultimate limitation regarding rejection rate comes from wavefront distortions, which result in a residual halo (volcano shaped) of unwanted light. This halo is not the trouble since it can be removed by an appropriate observing strategy. The trouble comes from shape fluctuation and photon noise, since they limit the detectability of faint companions. Numerical simulations (see [47]) based on a phase error map from HST have provided the profile of the expected halo at $\lambda = 2.2$ μm: a halo culminating at roughly 13 magnitudes below the maximum of the Airy profile has been found. Simulations also show that a SNR about 10 can be achieved in a 1000 sec integration for a magnitude difference ΔK = 14 if the central star has K = 8. This gives an encouraging starting point as soon as a comparable or even better optical quality will be achieved on NGST (for which the most limiting factor could actually be pointing instabilities). The range of magnitude differences and angular separations which should be reached allows good hopes concerning exoplanet detection.

Improvement: hybrid coronagraphy

In order to get rid from the 180° ambiguity in the resulting image, a focal mask can be inserted in the cat's eye focal plane (Baudoz *et al.* [6]). This hole will block all the light coming from the observed object except the central core of the diffraction pattern, so as to obtain a spatially filtered wavefront that is achromatically π-phase shifted. When recombining the two arms, we encounter the same interference process as in the case of a phase mask coronagraph. The only difference is that the π-phase shift is achromatic. A limitation still remains in the width of the focal mask that is optimized for a monochromatic light. When the bandwidth increases, the nulling efficiency decreases, but slower than when a phase mask is used. The diffraction effect from the focal mask can be reduced by using an apodized transmission of the pupil. Laboratory tests on a prototype have validated the concept [6].

2.3.5 Dark speckle technique

In the dark speckle technique (Labeyrie [25]), rapid random changes in optical path length due to atmospheric turbulence are exploited, with the goal of detecting the planet in very short exposures ($\sim$ 1 msec) when, by chance, the star light interferes destructively at the planet location. The planet's Airy peak, restored by an adaptive optics system, has a rather stable intensity, which adds to the star's local intensity. The intensity histogram is

therefore locally distorted, and suitable algorithms can display the local distortion in the form of a cleaned image. With a large telescope and hours of (discontinuous) integration, a planet 10^9 times fainter than its parent star is expected to become visible [11]. This technique can also be a precious complement to traditional coronagraphy techniques, as mentioned above. In both cases, the spatial and temporal resolutions of the camera have to be respectively better than the speckle area (about 500 pixel per speckle area) and shorter than their lifetime (typically about 10 msec). A narrow-band filter is also required because the diffraction and the speckle pattern are color-dependent.

Although it has been initially devised for ground-based telescopes, the dark speckle technique is also applicable to space telescopes, the seeing being provided by the deformable mirror reshaping. In that case, the individual exposure time is set by the adaptive optics stability, which is of the order of a few seconds in space. This considerably reduces photon noise compared to the ground-based observations.

2.3.6 Infrared nulling interferometry

Since angular resolution is the limiting factor when mid-IR observation is considered, infrared interferometry is an obvious solution for direct planet detection: the interferometer baseline can be adapted to any realistic angular separation. Moreover, the use of an interferometer can achieve starlight cancellation by means of destructive on-axis interference. This possibility has first been proposed by Bracewell [15] and will be described in full detail in the following chapter. We will prove that Earth-like planet detection and characterization by means of low-resolution spectroscopy is a reasonable goal with this technique.

2.3.7 Densified-pupil imaging interferometry

Fizeau interferometers, the equivalent of giant telescopes having a sparse mosaic mirror, produce an image, but its quality degrades catastrophically when the total aperture size becomes much larger than the sub-apertures (see section 1.2.1). A usable image can however be retrieved by densifying the exit pupil, i.e., distorting it to increase the relative size of the sub-pupils (see Labeyrie [26]). Such instruments, which may be called "hypertelescopes", evade a requirement long believed to be a golden rule of imaging interferometry, namely, that the exit pupil be identical to the entrance pupil. Instead, one preserves only the arrangement of the sub-pupil centers, while magnifying each sub-pupil with respect to the inter-pupil spacings, thus making the exit pupil more densely packed than the highly diluted entrance pupil (as explained in [13]).

In order to study circumstellar environments, it is possible to use a phase mask coronagraph on a densified-pupil interferometer (see Guyon and Roddier [20], Boccaletti *et al.* [13]). Pupil densification of a sparse array of apertures creates a PSF close to a single-aperture telescope PSF. Under those conditions, the phase mask coronagraph can work efficiently. Based on this principle, the proposed "Exo-Earth Discoverer" is a 36-element interferometer (each telescope about 0.6 m in diameter) with variable size ranging between 50 and 500 m, which has the short-term goal of obtaining unresolved images of exoplanets in the mid-IR range [13]. Simulations of a twin of the Solar System located at 20 pc have

shown that Earth-like planets can be detected in a 10-h exposure, even when dark-hole and dark-speckle techniques are not used. Note that an Achromatic Interfero-Coronagraph could be used instead of the phase-mask coronagraph. This possibility has not been evaluated yet.

Until now, we have only considered the goal of detecting Earth-like planets, without resolving their surface. The next step will be to obtain resolved images of such bodies, which requires interferometric arrays with a typical size of 100 km (and about 150 aperture elements of 3 meters). Densified-pupil interferometry can in principle provide such images, showing continental details and possibly "green spots" indicative of photosynthetic life (Labeyrie [27]).

2.4 Choice of the wavelength: a criterion for the existence of life

We have seen that four techniques are capable of detecting Earth-like exoplanets: photometric transits, gravitational microlensing, coronagraphy and nulling interferometry. Among those techniques, only the last two have the capability to characterize Earth-like planets by means of low-resolution spectroscopy. Mono-pupil coronagraphy is restricted to the visible and near-infrared spectral ranges because of angular resolution, whereas nulling interferometry with a typical 50 m baseline (and therefore better resolution) benefits from a lower contrast by observing in the mid-infrared range. A crucial advantage of the mid-infrared range is that the huge ratio of the star to planet fluxes is reduced by about three order of magnitudes with respect to the visible range. But recent breakthroughs in coronagraphic techniques have made it possible to fight against the huge contrast in the visible range.

The final choice between coronagraphy and nulling interferometry has therefore to take into account the capabilities of these two spectral ranges to evidence life by remote sensing, which is the ultimate goal of future planet finder missions as Darwin or TPF (see chapter 3 for a description of those missions). In this section, the most popular criterion for life is briefly described, as well as its consequences on the choice of the observation wavelength.

- *The H_2O - O_2 criterion.* Owen showed that, if based on chemistry, life is likely to rely on organic chemistry and to require the presence of liquid water (see Léger *et al.* [31], [32]). He also showed that if biological activity on a telluric planet develops on a large scale, it necessarily produces a large quantity of O_2. As this gas is highly reactive with reducing rocks and gases emitted by volcanoes, it would disappear in a short time in the absence of continuous production. The massive presence (1-1000 mbar) of O_2 would therefore be a strong indication of (primitive) life. O_2 has spectral signatures in the visible (two bands at 720 and 760 nm), whereas H_2O shows large absorption bands in the $6-8$ μm and $16-20$ μm spectral ranges. It is the reason why these two species cannot be simultaneously detected.

- *The H_2O - O_3 criterion.* An important step was made when Angel *et al.* [2] showed that ozone (O_3) is an attractive tracer of O_2. The spectral signature of ozone occurs

in the mid-infrared (9.6 μm), where the ratio of the star to planet fluxes is more favorable. Moreover, ozone has a logarithmic dependence upon the O_2 concentration: even a small amount of the latter would give rise to a significant O_3 band at 9.6 μm. The $6 - 18$ μm region seems therefore very informative, since it contains spectral features of H_2O and O_3, but also of CO_2 at 15 μm (see figure 2.10). The raw materials in the primitive atmospheres of terrestrial planets are indeed expected to be fully oxidized carbon (CO_2). Finally, the detection of CH_4 at 7.7 μm, which requires a higher spectral resolution (around 1000), would strongly suggest biological activity as we know it on Earth.

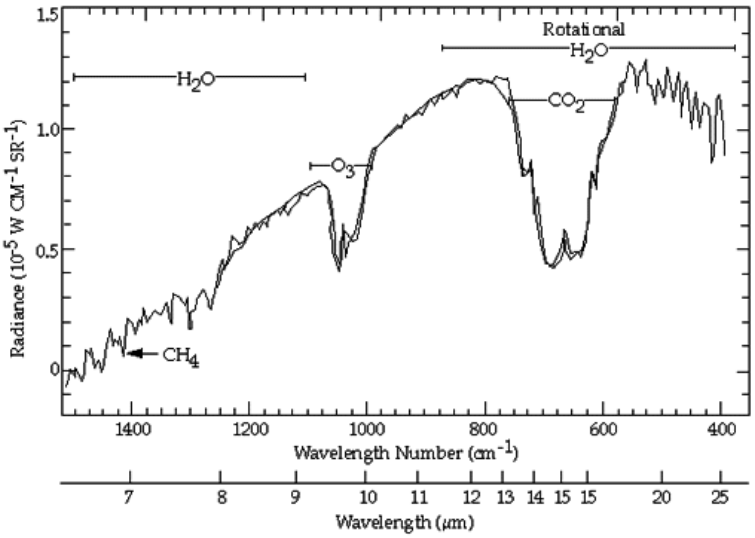


Figure 2.10: The infrared spectrum of the Earth as observed from an orbiting Nimbus satellite shows a variety of spectral features. Taken from the ExNPS Road Map web-site.

For all these reasons, the current baseline for future exo-Earth finder missions (Darwin and TPF) is a nulling interferometer working in the $6 - 18$ μm spectral range. In order to safely detect the ozone band, a resolution of 20 should be reached. The Darwin concept seems to be our best hope to achieve these performances, and thus to detect habitable planets outside our Solar System for the first time.

Chapter 3

Nulling interferometry and ESA's DARWIN mission

In this chapter, we focus on space infrared interferometry, and especially on ESA's DARWIN mission, whose main goal is to detect and characterize terrestrial exoplanets. Unless explicitly specified, a typical target consisting in a Sun-Earth system at 10 pc is assumed. Since the orbital plane is randomly oriented, a mean angular separation of about 80 mas is assumed (see [39]).

3.1 Principle of a Bracewell interferometer

As we have seen in the last chapter, there are two major problems to be solved to directly detect exoplanets: the high contrast requires a broad dynamic range, and on the other hand, a high angular resolution is necessary. However, while the angular separations are fixed constraints, the contrast ratio is amenable to modification in at least two ways: observing in the infrared and altering the intrinsic contrast ratio with an optical device. Both of these have been used by Bracewell to design his famous nulling interferometer ([15], [16]).

3.1.1 Description

The basic principle of nulling interferometry is quite simple: combine the light incident on a pair of telescopes in a co-axial mode, adjusting their respective phase in order to produce a totally destructive interference on the optical axis. The two electric fields must be π radians out of phase to achieve that kind of interference.

Figure 3.1 illustrates the theoretical design of such an interferometer, where the recombining system induces an achromatic[1] phase shift of π radians. If θ represents the angular distance from the axis to the star projected along the interferometer baseline, the interference pattern resulting from the superposition of the two waves is proportional to $\sin^2(\pi\theta/\Phi)$, if the observed quantity is an intensity (i.e., the square of the complex amplitude). The parameter $\Phi = \lambda/L$ is the angular separation of the first out-of-axis minimum,

[1] i.e., independent of wavelength

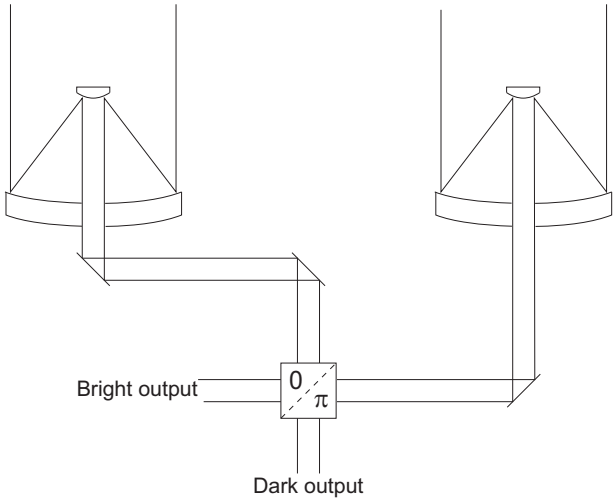


Figure 3.1: Sketch of Bracewell's nulling interferometer. The recombining system provides an achromatic π phase shift on one of the two outputs.

with L the distance between the two telescopes and λ the observation wavelength. It depends on the distance between the two telescopes (or baseline) and on the observation wavelength. The interference pattern, projected on the plane of the sky, shows alternate dark and bright fringes (figure 3.2), corresponding to the minima and maxima of the sinusoid. It modulates the planetary signal by a factor ranging between zero and one. The configuration of the interferometer is chosen so that the interference pattern $\sin^2(\pi\theta/\Phi)$ has a maximum on a planet at $\theta = \Phi/2$ when response at the star ($\theta = 0$) is minimized.

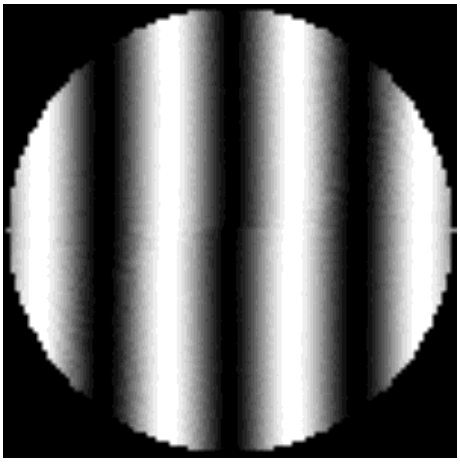


Figure 3.2: Transmission map for a perfect Bracewell nulling interferometer with $L = 25$ m and $\lambda = 10$ μm over a 500×500 mas field of view.

Since the only information we get while observing the planetary system is an integrated flux, we cannot distinguish between a bright source on a dark fringe and a faint source on

a bright fringe. Moreover, the faint planetary signal has to be detected in the presence of unwanted signals (see section 3.1.2). That is the reason why Bracewell proposed to continuously rotate the interferometer with periods of about an hour to modulate the planetary signal, the planet alternately crossing dark and bright fringes. The signal modulation of an off-axis point-like source is strongly dependent on two parameters:

- The coordinates of the source relative to the central star. The further the planet, the more fringes it crosses during a complete array rotation, and so the highest the modulation frequency.
- The observation wavelength and the baseline. Indeed, the interferometric pattern varies with both of them: the pattern is narrower when the wavelength is shorter or the baseline longer.

Therefore, a simple Fourier analysis will be sufficient to localize orbiting planets. For example (figure 3.3), if the planet has $\theta = \Phi/2$, then the signal will vary as $\sin^2[(\pi/2)\cos\omega t]$, where ω represents the frequency of the interferometer spin. This function has a fundamental frequency of 2ω, but also contains a noticeable amount of the 4ω harmonic. Very faint signals can be recovered by synchronous detection if they are modulated at a known frequency (Bracewell and MacPhie [16]).

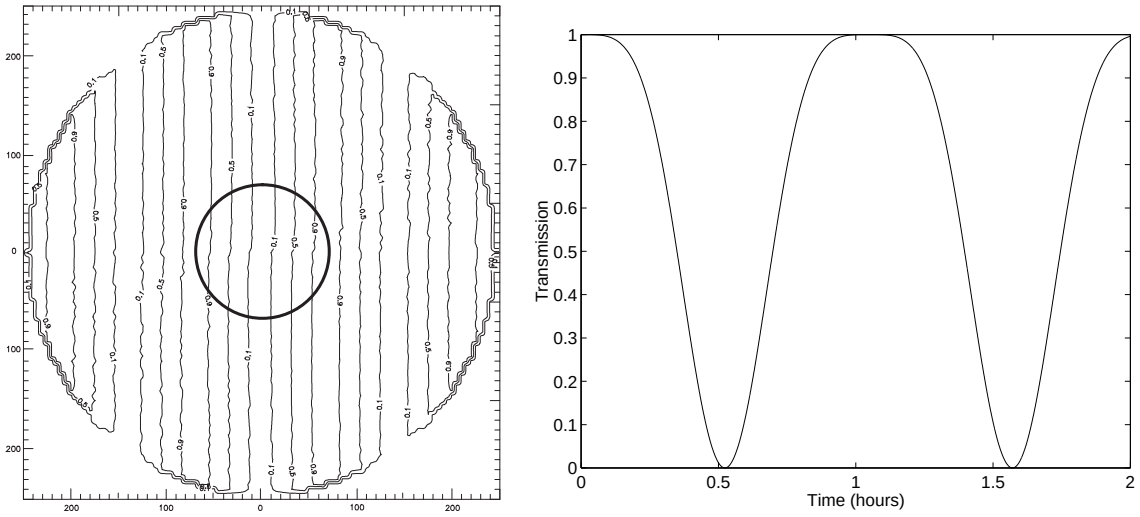


Figure 3.3: Left: path of the planet on the transmission map for a perfect Bracewell interferometer, with the planet at $\theta = \Phi/2$. Right: modulation of the planetary signal, at a principal frequency $2\omega = 2.8 \times 10^{-4}$ Hz (one rotation completed every two hours).

The essential characteristic of such an interferometer is its high dynamics, which is measured by the *rejection rate* ρ, defined as the ratio of the stellar flux that would be observed without a π phase shift over the one observed after the destructive interference. Note that this ratio varies as θ^{-2} for small θ, because of the $\sin^2(\pi\theta/\Phi)$ dependence of the pattern for a Bracewell interferometer.

3.1.2 Fundamental limitations

Bracewell's concept, however interesting, will not be used as such to detect Earth-like planets, even with a space interferometer. Its limitations are related to extraneous signals that add to the planetary signal, and to the incapacity to remove them in order to extract the useful information.

Stellar leaks

Even with an ideal Bracewell interferometer, the stellar signal is only perfectly cancelled on the optical axis. The finite stellar diameter allows for light from the edges of the star to "leak" through the fringe pattern. As we will see in the next section, the theoretical rejection rate is related to the spatial configuration of the entrance pupil. In the case of a two-telescope nulling interferometer, the rejection rate grows with decreasing angular distance in a $\rho \propto \theta^{-2}$ law, which gives typical rejection rates of order 10^3 for nearby main sequence stars (see section 3.4.2). This rate allows the detection of hot Jupiters, but Earth-like planets are completely out of reach.

Figure 3.4 shows a cross-sectional view through the null fringe and the star (thick line above the horizontal axis) for three rejection functions. It illustrates the leakage of the star edges through the fringe pattern.

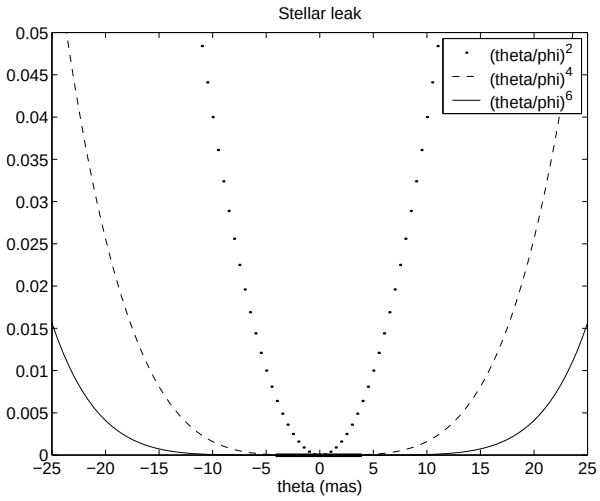


Figure 3.4: Stellar leak for $\Phi/2 = 50$ mas (position of the first maximum). The three curves illustrate the θ^2, θ^4 and θ^6 dependence of the starlight transmission close to the axis. The bold horizontal line represents the size of the stellar disc.

The noise created by the star can be separated into two components: the constant component resulting from the incomplete cancellation of the stellar flux, and the variable component resulting from the variation of the rejection rate, due to optical path difference (OPD) and differential pointing variations. This last term, called "leakage jitter", can be shown to be negligible for jitter values < 1 mas (see [8]).

Exo-zodiacal light

Besides proximity and contrast, a third obstacle to direct exoplanet detection is the possible emission from exo-zodiacal dust grains congregating in and near the orbital plane of the target planetary system. Such dust disks can be expected both by analogy with our own solar system's zodiacal cloud, and by extrapolation of the few known cases of very bright circumstellar disk emission.

The disk's morphology and intensity are a priori unknown. Suppose it has a circular symmetry, and the same intensity as the solar zodiacal cloud, we can then infer some characteristics of the resulting modulated signal. If the system is seen pole-on, the exo-zodiacal signal is not modulated by rotation of the array, and the planetary signal can be easily isolated. On the contrary, if the line-of-sight is not perpendicular to the ecliptic plane of the exo-system, circular symmetry won't be observed anymore, and the centro-symmetric exo-zodiacal signal will be modulated, at the same rate as the planetary signal since the array has central symmetry.

The problem is all the more serious because the total integrated flux from the exo-zodi is 400 times higher than an earth-like planet flux, when observed at 10 μm. Taking into account the multiplication by the transmission map that cancels the center of the exo-zodiacal light, this ratio reduces by a factor $\simeq 3$, but the problem still remains (see figure 3.5). This source also contributes to the background radiation, and its noise has to be taken into account.

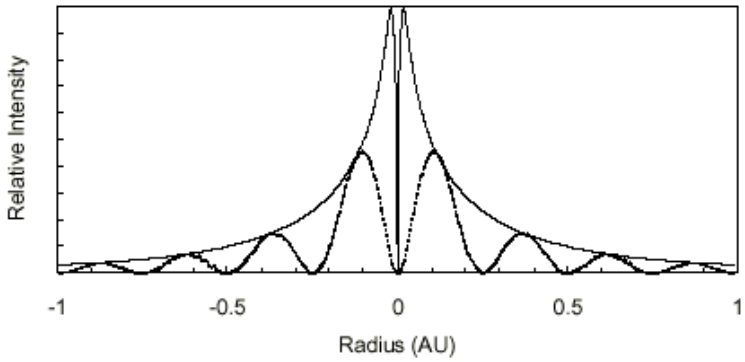


Figure 3.5: Cross-sectional view through an edge-on exo-zodiacal cloud at distance 10 pc, showing the emergent flux and the flux transmitted by the 10 μm fringe pattern of the Keck interferometer in the nulling mode (from [28]).

In order to scale the interferometer for adequate sensitivity (and particularly if the exo-zodiacal light is the main source of noise), the 10 μm emission from such dust clouds must be characterized prior to launch of the space interferometer by means of a ground-based interferometer or coronagraph (see chapter 5).

The ability to identify an Earth-like planet would also be severely degraded if there were "random" exo-zodiacal brightness variations with 10% amplitude. Fortunately, the overall zodiacal cloud is expected to be relatively homogeneous because of the smoothing effect of the dissipation forces, and therefore, inhomogeneities should not affect planet detection.

The background: local zodiacal light

One of the main noise sources of background radiation is the thermal emission due to dust accumulated in the local zodiacal cloud. It creates a diffuse glow at infrared wavelengths that hampers the view of the sky, just as city's lights hide the faintest stars from our gaze. The radiation gets brighter and "warmer" as we approach the Sun, and is also dependent on the line-of-sight. For a near-Earth orbit, and a line-of-sight at 90° from the Sun direction in the ecliptic plane, the emission can be represented by a grey body emission at a temperature of 235 K and with a grey emissivity of 3.8×10^{-7} (see [33], p.72).

Being an extended source, the zodiacal light gives a total flux proportional to $S\Omega = 3.7\,\lambda^2$ (see [54]) where S is the telescope surface and Ω the solid angle defined by the first dark ring of the Airy pattern at the detector. On the other hand, the planetary signal is proportional to the collecting area. If the interferometer orbits at 1 AU from the sun with 1 meter class telescopes, the integrated zodiacal light is typically 10^3 times larger than an Earth-like planet signal at 10 pc, but is not modulated by rotation of the array. The drawback of such an emission is twofold:

- The noise generated by the local zodi can prevent from detecting the planetary signal.
- The long term instability of the background over typical periods of an hour introduces another difficulty in the detection of the planet signal, since it induces signal fluctuations at the typical modulation frequency of the planet signal.

Since the collected zodiacal light is independent of the diameter, the signal-to-noise ratio varies as D^2 (D representing the diameter of the apertures) and the integration time as $1/D^4$. It is the reason why the first proposed configurations had very large apertures. A detailed study of the zodiacal cloud emission can be found in Eric Thomas' thesis [54].

Other contributors to the background

The thermal background emission of the interferometer can generate troublesome levels of background signal whenever the temperature of the cold optics exceeds 40 K (see [34]). Another contributor to the background is the detector dark current. It should be below 10 e^-/s/pixel per spectral channel in order to remain negligible when compared to the local zodiacal light background. The thermal emission of the interferometer should also be very stable in time, as well as the efficiency of the detectors, the gain of the amplifiers, ...

One of the principal advantages of space observation is that the optics can be cooled to reduce their proper infrared emission. The goal is to achieve a thermal background smaller than the contribution of the local zodiacal light. This is the case if the optics are cooled under 40 K, which can be reached with passive cooling.

3.2 Improvements of the basic concept

Since 1979, many improvements of Bracewell's original concept have been proposed. In this section, I will summarize the chronology of the main ideas, emphasizing on the theoretical aspects needed to proceed in this work.

3.2.1 Central attenuation

The interferometer response

This paragraph is largely based on the work of Mennesson and Mariotti ([37], [40]).

As stated in the last section, the most important source of noise comes from the stellar leaks in Bracewell's configuration, whose rejection rate is smaller than the required rate[2] by about three orders of magnitude. To obtain a rejection rate $\rho \geq 10^6$, the transmission of the interferometer near the optical axis must be proportional to θ^4 or even better to θ^6 (see figure 3.4), where θ is the off-axis angle when pointing towards the star. In order to achieve that functional dependence, more than two independent apertures are needed.

Consider the case of an interferometer with n telescopes in a two-dimensional array (figure 3.6). The telescope coordinates are characterized by polar coordinates (L_k, δ_k) defined with respect to the center of the array, taking one of the interferometer arms as a reference. The diameter of each telescope is denoted D_k. We assume that an arbitrary achromatic phase shift ϕ_k is applied to each beam before recombination. The location of a point source is defined by the angular separation from the star (θ) and the azimuth relative to the first interferometric arm (ϕ).

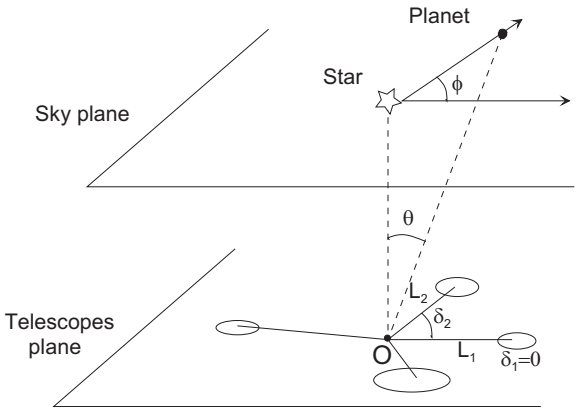


Figure 3.6: Geometrical configuration of the array and the stellar system. Planetary coordinates are given by the off-axis angular direction $\vec{\theta} = (\theta, \phi)$. Each telescope of index k is located by means of the vector $\vec{L_k} = (L_k, \delta_k)$ in a reference frame with arbitrary origin O, used as a common phase reference.

Neglecting aberrations, the detected complex amplitude from a point-like source $\vec{\theta} = (\theta, \phi)$ is given by (see Mennesson's thesis [37], p.136)

$$\mathbf{A}(\vec{\theta}, \vec{r}) = \Pi(r/R) \left(\mathbf{A}_1(\vec{\theta}) e^{j\phi_1} + \ldots + \mathbf{A}_n(\vec{\theta}) e^{j\phi_n} \right) , \tag{3.1}$$

where $\vec{r}$ is the coordinate on the circular output pupil (assumed to be circular of radius R), Π is a rectangular box function[3] and where the moduli A_k of the complex amplitudes $\mathbf{A}_k$

[2]The rejection rate needed to detect Earth-like exoplanets within a reasonable integration time is larger than $\sim 10^6$ (Léger *et al.* [31]).

[3]$\Pi(r/R) = 1$ if and only if $r \in [0, R]$, with $r \geq 0$ by definition.

are proportional to the telescope diameters D_k, provided that the same transmissions affect the n beams. The complex amplitudes also depend on the external OPD with respect to the origin O of the axes: $\vec{L_k} \cdot \vec{\theta} = L_k \theta \cos(\delta_k - \phi)$, giving the expression

$$\mathbf{A}(\vec{\theta}, \vec{r}) = \Pi(r/R)\, e^{j2\pi \vec{r}\cdot\vec{\theta}/\lambda} \sum_{k=1}^{n} A_k e^{j2\pi (L_k\theta/\lambda)\cos(\delta_k - \phi)}\, e^{j\phi_k} \,. \tag{3.2}$$

The amplitude distribution in the focal plane is obtained by Fourier transform, which will not alter the constant terms following the summation sign. Denoting θ_f the angular coordinate in the focal plane and $\otimes$ the convolution sign, we obtain:

$$\widetilde{\mathbf{A}}(\vec{\theta}, \vec{\theta_f}) \propto \frac{2J_1(\pi D\theta/\lambda)}{\pi D\theta/\lambda} \otimes \delta(|\vec{\theta_f} - \vec{\theta}|) \sum_{k=1}^{n} A_k e^{j2\pi (L_k\theta/\lambda)\cos(\delta_k - \phi)}\, e^{j\phi_k} \,. \tag{3.3}$$

Thus, the intensity distribution $I(\vec{\theta}, \vec{\theta_f}) = |\widetilde{\mathbf{A}}(\vec{\theta}, \vec{\theta_f})|^2$ in the focal plane is the product of a classical diffraction term and a constant interference term, which only depends on the aperture configuration. Using a monopixel detector and neglecting diffraction features, we obtain a signal of energy proportional to the *interferometer response*:

$$R(\theta, \phi) = \left| \sum_{k=1}^{n} A_k e^{j2\pi (L_k\theta/\lambda)\cos(\delta_k - \phi)}\, e^{j\phi_k} \right|^2 \,. \tag{3.4}$$

Generalized entrance pupil

As stated by Mennesson [38], the generalized entrance pupil[4] P in the telescope plane of coordinate $\vec{r}$ is defined by:

$$\begin{cases} P(\vec{r}/\lambda) = A_k e^{j\phi_k} & \text{if } \vec{r} \text{ is in the aperture of telescope } k\,, \\ P(\vec{r}/\lambda) = 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

For point-like telescopes located at $\vec{L_k}$, it can be rewritten in our case as:

$$P(\vec{r}/\lambda) = \sum_{k=1}^{n} \delta(\vec{r}/\lambda - \vec{L_k}/\lambda) A_k e^{j\phi_k} \,. \tag{3.6}$$

Note that the generalized entrance pupil is complex in the general case, and that the interferometer response $R(\theta, \phi)$ is the square modulus of the Fourier transform of $P(\vec{r}/\lambda)$.

Conditions for a θ^{2p} transmission

Let x_k be equal to $2\pi(L_k\theta/\lambda)$. Close to the line-of-sight, we can develop the interferometric response in a power series

$$\begin{aligned} R(\theta, \phi) &= \left| \sum_{k=1}^{n} A_k\, e^{j\phi_k} \left(1 + j x_k \cos(\delta_k - \phi) - (x_k^2/2)\cos^2(\delta_k - \phi) + \dots\right) \right|^2 \\ &= \left| \sum_{k=1}^{n} A_k\, e^{j\phi_k} + \sum_{k=1}^{n} A_k\, e^{j\phi_k} j x_k \cos(\delta_k - \phi) - \dots \right|^2 \,. \end{aligned} \tag{3.7}$$

[4] This concept is derived from the classical definition of the entrance pupil of an optical system, see [33].

To have a θ^{2p} dependence for the stellar leaks, the first $p-1$ coefficients must cancel out, so that $x_k^p \propto \theta^p$ is the first contributor to the sum. Assuming $A_k \propto D_k$ (same transmission for the n light beams), the cancellation of the first $p-1$ coefficients yields the condition:

$$\sum_{k=1}^{n} D_k \left(L_k \cos(\delta_k - \phi)\right)^q e^{j\phi_k} = 0 \ , \ \forall q, \ 0 \leq q \leq p-1 \text{ and } \forall \phi \ . \tag{3.8}$$

Now, we can prove mathematically that the rejection rate in Bracewell's configuration (two telescopes) has a θ^{-2} dependence. Indeed, we immediately see that the above condition is only satisfied for $q = 0$, leading to $p = 1$ and a θ^2 dependence for the central transmission.

3.2.2 A better rejection rate: the Angel Cross and its derivatives

Chronologically, the second proposed nulling configuration is the Angel Cross (proposed by J. R. Angel in 1990). This configuration makes use of four telescopes in a square shape, combined with multiple beam-splitters for transmission varying as $\sin^4\theta$ instead of $\sin^2\theta$. The Angel Cross is indeed a double Bracewell interferometer, where the two outputs are recombined with no additional phase shift (figure 3.7). Angel's first idea was to use four 8 m-class telescopes standing at the summits of a square about 30 m wide, located at 1 AU from the Sun and achieving an Earth-like planet detection at 10 pc in a few hours of integration.

The original Angel Cross configuration has been further investigated in the past few years, giving rise to new possibilities with a θ^{-4} rejection rate, which are presented below.

Degenerated Angel Cross (DAC)

The Degenerated Angel Cross (DAC) consists of three telescopes in a linear array, with the central telescope's diameter two times larger than the two others and π radians out of phase (figure 3.7). Its name stems from the squeezing of an Angel Cross along one arm. This configuration satisfies the condition (3.8) for $p = 2$, yielding a θ^{-4} rejection rate. It can be shown (see appendix C.1) that it is not the only 3-telescope linear array with a θ^{-4} rejection: such a configuration exists whatever the distances between the three telescopes, but the outer telescopes does not have the same diameter any longer (the relation $D_1 L_1 = D_3 L_3$ must be verified). This set of configurations is referred hereafter as to the DAC family. The regular DAC has a particular status in this family, because of its symmetry. In the following, when speaking of DACs, the use of a regular DAC will generally be assumed.

When looking closer at the beam-combination scheme (figure 3.8), one can see that the definition of a DAC has to be slightly modified because, in the recombination scheme, the amplitudes A_k are not exactly proportional to D_k. Indeed, the light from telescopes 1 and 3 crosses two beam-combiners, whereas the light from telescope 2 only crosses one. For each beam-combiner, there are two outputs: one of them is constructive, and the other destructive. Each output carries one half of the total energy (i.e., $1/\sqrt{2}$ times the initial amplitude). So, the two light beams entering the second beam-combiner do not have the same amplitude, and do not cancel out for an on-axis source as desired, since the resulting amplitude is $A \propto 1/\sqrt{2}(1/\sqrt{2} + 1/\sqrt{2} - 2) \neq 0$. Consequently, if one wishes to use a single DAC, the diameter ratio must be $1{:}\sqrt{2}{:}1$, and not 1:2:1.

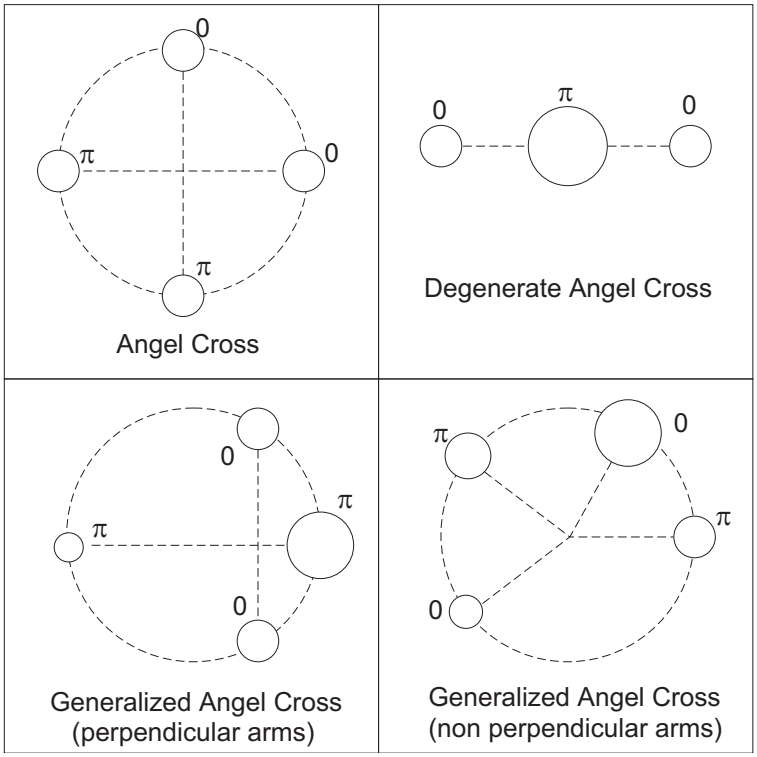


Figure 3.7: The Angel Cross configuration and its derivatives. The value near each telescope is the phase shift applied at the recombination.

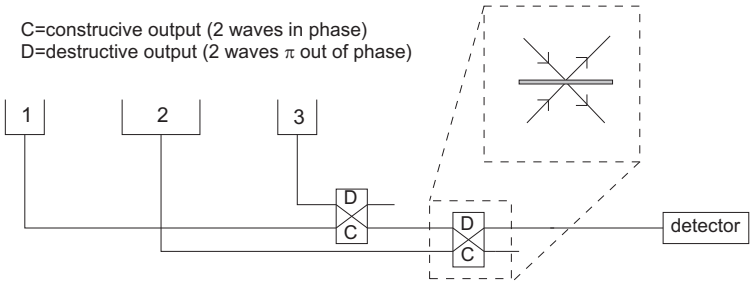


Figure 3.8: Recombination scheme for the regular DAC. This scheme proves that this DAC cannot be used with diameters 1:2:1 unless the intensity of the light beam coming from telescope 2 is reduced (see text). Note that one half of the intensity is lost at each beam-combiner, since only one of the two outputs is exploited.

Generalized Angel Cross (GAC)

The Generalized Angel Cross (GAC) represents any 4-telescope nulling interferometer with a θ^{-4} rejection rate restricted to achromatic phase shifts of 0 or π radians. In general, the telescopes are of different sizes. To avoid long delay lines, the four telescopes must be

located at equal distances from the beam-combiner in a common plane[5], i.e., on a circle with the beam-combiner at its center. In short, the GAC configurations are the plane 4-telescope configurations assuming constant distance between the beam-combiner and the telescopes, satisfying the conditions for a θ^{-4} rejection:

$$\sum_{k=1}^{n} D_k \, e^{j\phi_k} = 0 \tag{3.9}$$

$$\sum_{k=1}^{n} D_k \, \cos(\delta_k - \phi) \, e^{j\phi_k} = 0 \, , \text{ for all } \phi. \tag{3.10}$$

Figure 3.7 shows that GACs can have perpendicular or non perpendicular arms. In the last case, it doesn't any longer look like a cross.

Note that GACs do not suffer from the same recombination problems as DACs, because each light beam crosses two beam-combiners on its path to the detector. The only trouble is that only a quarter of the total intensity reaches the detector, because one half is lost in each recombination process.

The first Darwin configuration

The first concept for Darwin, proposed by Léger, Mariotti *et al.* in 1993, was based on an Angel Cross, but with four 1.5 m class telescopes orbiting at 4 or 5 AU. Telescopes of that size seem more reasonable than Angel's proposal (with 8 m-class telescopes), and compatible with an Ariane 5 shroud. Small telescopes can be used at 5 AU because the surface brightness of the zodiacal light drops by a factor 1800 between 1 and 5 AU (see figure 3.9). At that distance, the (still unknown) exo-zodiacal light becomes the most important source of noise. In 1996, a first quantitative study of this Darwin configuration was published by Léger *et al.* [31]. Since then, the Darwin concept has been significantly improved (as will be described in the next sections), and was finally selected by the European Space Agency (ESA) as a candidate cornerstone for the Horizon 2000+ program (October 2000).

3.2.3 Asymmetric arrays

The main trouble with the Bracewell and Angel Cross arrays is their central symmetry. As explained in section 3.1.2, that symmetry prevents the detection of Earth-like planets because the bright exo-zodiacal light is modulated at the same frequency as the planetary signal (except if the exo-zodi is seen pole-on, in which case it is not modulated at all). The solution is to design an interferometer with an asymmetric transmission map, so that every planetary system can be treated the same way, whatever its inclination.

Five telescopes on a circle

In their 1996 paper, Léger *et al.* [31] proposed an asymmetric array, made up of five telescopes, regularly located on a circle, with a θ^{-4} rejection law. This configuration

[5] We will see later (section 3.3.1) that a plane configuration is needed to avoid thermal coupling between the telescopes.

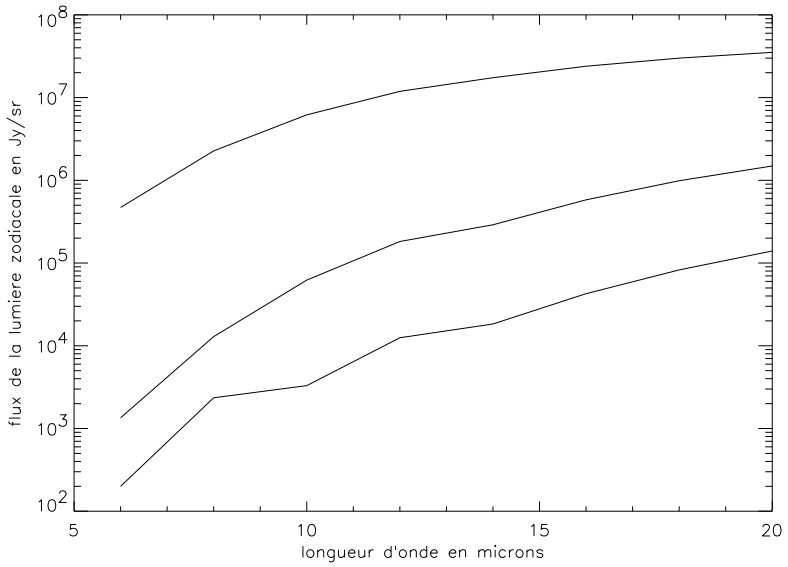


Figure 3.9: Local zodiacal cloud brightness as a function of wavelength at three distances from the Sun: 1 AU (top curve), 3 AU (middle curve) and 5 AU (bottom curve). These curves are relevant for Darwin's typical observing directions. Figure from B. Mennesson [37].

has been further studied in a paper by Mennesson and Mariotti (1997, [40]), on which this paragraph is largely based. To get a θ^{-4} rejection rate, the five telescopes must be regularly spaced on the circle, and the phase shifts equal to $\phi_k = 4(k-1)\pi/5$, which implies a complex generalized entrance pupil. The transmission map of the interferometer, projected on the plane of the sky, is shown in figure 3.10.

If the array is continuously rotated at a frequency $f_0 \sim 1$ turn per hour, the signal emitted by a planet located on the first set of bright fringes will mostly exhibit a $5f_0$ frequency, whereas the exo-zodiacal signal will concentrate at $10f_0$ and higher even harmonics, because of its central symmetry. This fact allows to extract the planetary contribution to the modulated signal. But two major problems remain with the circle configuration:

- The planet's image and location will only be known modulo $2\pi/5$, leading to five possible locations for the planet. This undetermination stems from the fact that the transmission map shows a five-fold symmetry.

- The transmission map shows a big crown-shaped hole of transmission, and thus a region of very weak modulation, that makes it necessary to change the baseline in orbit. This is particularly important for spectroscopy, to get a more uniform spectral coverage since the scale factor of the transmission map is proportional to the wavelength λ.

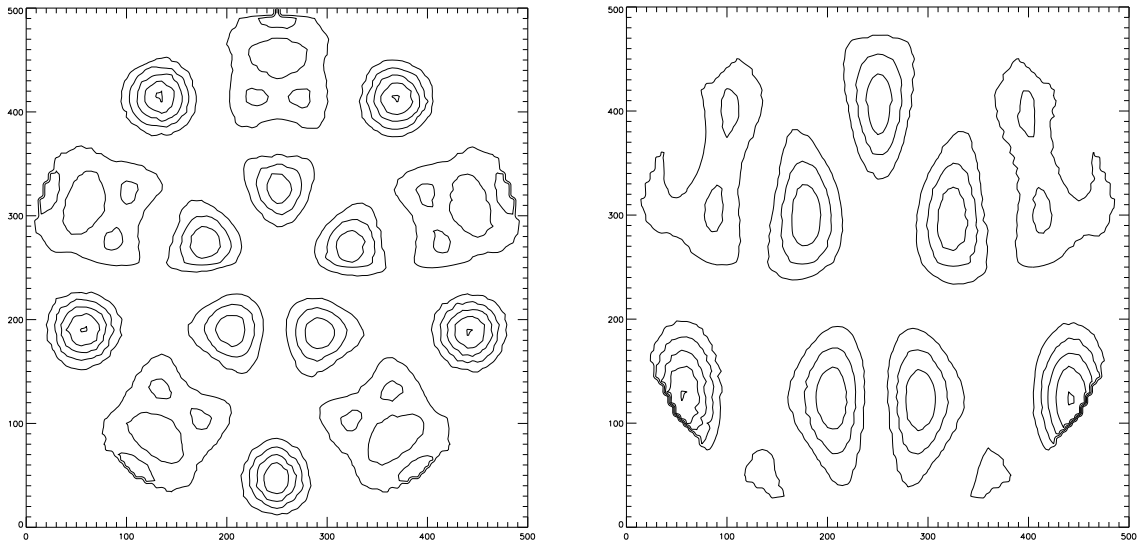


Figure 3.10: Transmission map of the circular array (left) and of the elliptical array (right) projected on the sky. The central symmetry of the map is broken, leading to easy distinction between the signals emitted by an exo-planet and by the exo-zodiacal light. Transmission curves are drawn for transmission ranging from 20% to 100% of maximum (coordinates in mas, observation wavelength 10 μm and array radius of 15 m).

Five telescopes on an ellipse

In the same article, the authors propose an improvement of the circular array, based on an elliptical configuration. In fact, any configuration keeping the same ϕ_k and deduced from the circle by a geometric affinity (e.g. by contraction along any direction) still works.

The elliptical configuration shows an irregular transmission map that permits the exact localization of the planet, since we can recognize between the five maxima in the transmission map (see figure 3.10). Moreover, azimuthal holes are not present any more if the array is rotated around its center, and every angular separation has a good amplitude of modulation[6]. Thus there is no more need for a variable baseline.

But a new problem arises with this configuration: since the telescopes are situated at different distances from the recombining hub (at the center of the ellipse), continuously variable long delay lines are required to equalize the optical path lengths of the incoming beams. The requirement of such delay lines is considered to add considerable additional complexity to the system, and should be avoided unless absolutely necessary. On the other hand, short active delay lines would probably be needed for all concepts, to facilitate the search of the fringes on the science object and to compensate for internal (mostly thermo-elastic) biases (see [34]). Unlike long variable delay lines, short ones do not constitute a major difficulty.

[6]The amplitude of modulation is measured by the root mean square of the transmission over a rotation period, see [40] for details.

General conditions to get an asymmetric interferometer response

As we have just seen, non-real entrance pupils seem to give transmission maps without central symmetry. In appendix A, we prove[7] that, for configurations with more than two telescopes:

- Real generalized pupils always lead to centro-symmetric transmission maps.
- As expected, non-real pupils generally lead to non-symmetric maps. However, in several particular cases, they can also yield centro-symmetric transmission maps, but these special configurations seem inappropriate for nulling (see appendix A).

This is an important result, which will be used many times in the following developments.

Other important information can be deduced from analytical developments: if the telescopes are restricted to lie on a circle (in order to avoid long delay lines), and if a θ^4 starlight suppression is needed, then an asymmetric transmission map cannot be obtained with less than five telescopes (see appendix B). That is the reason why asymmetric transmission maps with 3 or 4 telescopes on a circle have never been found. It also means that, besides linear arrays and other special configurations where long delay lines could be avoided, five telescopes are needed to perform a secure detection of Earth-like planets with a nulling interferometer. We will later see that this statement remains almost true (for other reasons) when using internal modulation.

3.2.4 The need for chopping

Besides rotation of the array, which provides slow modulation of the planetary signal versus other contributors, we shall use an additional level of modulation, a technique called "chopping", to modulate the desired planetary signal in presence of varying backgrounds and detector drifts. As already mentioned in section 3.1.2, long term instability of the background over typical periods of an hour reduces the interferometer's sensitivity, unless the rotation period of the array is much smaller than the typical timescale of background drifts.

If the desired signal can deliberately be modulated (chopped) at a frequency that no other signal is likely to mimic, then a chopped planet signal will systematically average to a steady value while the drifts will tend to average to zero. In practice, the chopping frequency is set by the minimum readout time of the detector, such that the readout noise remains much smaller than the detected photon flux. Frequencies are therefore generally of about 0.1 to 0.01 Hz. This will also greatly relax the constraint on the rotation frequency of the interferometer.

Chopping has long been used in infrared and radio ground-based astronomy. The principle of the technique is to rapidly alternate between the observations of the stellar source and of the background. It is well suited to image objects fainter than the thermal background. The trick is to point alternately the star and a nearby sky position, by periodically tilting one of the mirrors (generally the secondary or the tertiary mirror). The background can then be properly subtracted from the scientific data.

[7] This demonstration is inspired from a fruitful discussion with Bertrand Mennesson.

In the case of a nulling interferometer, if we cease to point the star, the central dark fringe will not occult the star any more, and a large amount of starlight will therefore be detected. An easy measurement of the background cannot be achieved the same way as in classical infrared observations.

However there are a number of configurations for an interferometer that can be used to provide both nulling and chopping. The principle is to use two (or more) subsets of nulling interferometers, and combine them is such a way that the point-like planetary signal is strongly modulated whereas the extended background sources are not. In the following discussion, we will focus on a particular way to achieve both nulling and chopping: the so-called internal modulation.

3.2.5 Internal modulation

Principle of internal modulation

We have discussed two ways to improve the extraction of the planetary signal from other unwanted sources: external modulation (rotation of the whole array) performed with an asymmetric transmission map[8], and chopping. However, "classical" chopping does not seem appropriate for nulling interferometry and until 1997, the only modulation level was physical rotation of the array, with two main drawbacks:

- Reduction of the observation time, since a lot of time is spent in rotating the array.
- Impossibility to modulate the planet signal at a high frequency, which is necessary to compensate for fluctuations in background levels.

To overcome those main problems, we must possibly avoid any physical rotation of the array. That is the purpose of internal modulation. The basic idea is to apply a variable phase shift between the nulled outputs of two interferometric sub-arrays (e.g. Bracewell, Angel Cross, DAC, GAC, ...), recombine them and make the detection. This modulation is called internal because it only takes place at the recombination of the beams. Although double arrays[9] were already considered by Woolf and Angel [59], the principle of internal modulation and its practical implementation were proposed by Jean-Marie Mariotti[10] in 1997, and published in 2000 by Mennesson and Léger [39].

An analysis based on Fresnel complex amplitude vectors, proposed by Mennesson and Léger [39], shows the effect of such a variable phase shift $\phi_{im}(t)$ for the case of two Bracewell sub-arrays (see figure 3.11). When the two outputs are nulled (e.g. for an on-axis source), the total output is also nulled, whatever the value of $\phi_{im}(t)$. But when the two outputs are constructive (e.g. for a well located planet), the total output varies between zero (for $\phi_{im} = \pi$) and a maximal value equal to the sum of the two outputs (for $\phi_{im} = 0$).

[8]Note that a necessary condition to get an asymmetric map is that some of the phase shifts ϕ_k are different from 0 or π, which is still a technical challenge (see section 3.4.4).

[9]Double arrays consist in two paired linear arrays close together with their elements interdigitated.

[10]Jean-Marie Mariotti suddenly died in July 1998. He was deeply involved in the space exo-biology missions currently under study: Darwin in Europe and TPF in the United States.

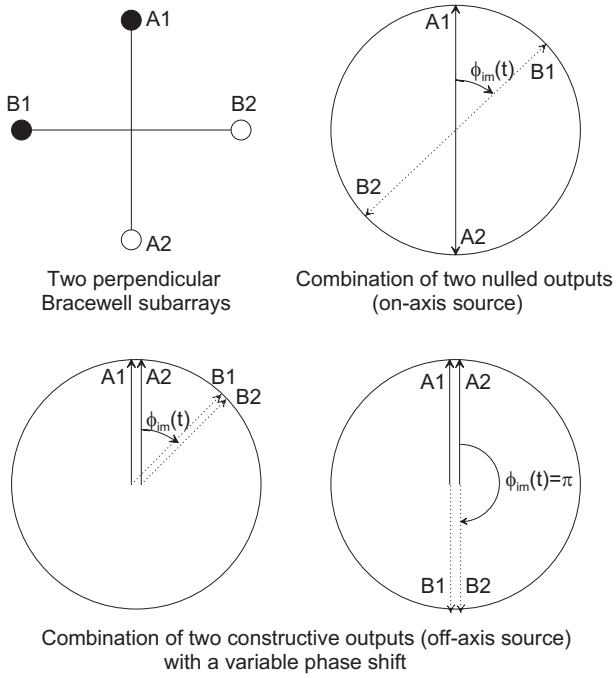


Figure 3.11: Principle of internal modulation illustrated with two Bracewell sub-arrays (from Mennesson [37]). For an on-axis source, both Bracewell outputs are null, so that the combination of these is also null and not modulated. On the other hand, an off-axis source located on a constructive fringe for both Bracewell interferometers is strongly modulated, with the same period as the variable phase shift $\phi_{im}(t)$.

Thus, the signal of an off-axis planet is internally modulated as $\phi_{im}(t)$ varies, with a modulation amplitude depending on the planet's projected location. Since this process can be done at high frequency (typically 0.1 Hz), it achieves both modulation and chopping in one single process, without any physical rotation of the array (note that in practice, several small rotations will anyhow be required).

Recombination of two sub-arrays

The two following paragraphs are based on the work of Mennesson ([37], [38]).

We consider here the recombination of two nulling interferometers with a time variable phase shift $\phi_{im}(t)$. Let $\mathbf{A}_1(\vec{\theta})$ and $\mathbf{A}_2(\vec{\theta})$ be the complex amplitudes of the electric field transmitted by each of the two sub-nulling interferometers, originating from a common point source located at an off-axis position $\vec{\theta}$ from the star. The intensity detected on one side of the beam-combiner is proportional to:

$$S_{12}(\vec{\theta}) = \frac{1}{2}|\mathbf{A}_1(\vec{\theta}) + e^{j\phi_{im}(t)}\mathbf{A}_2(\vec{\theta})|^2 \,, \tag{3.11}$$

where the $1/2$ factor comes from the equal division of the signal between the two outputs of the beam-combiner. We will always assume beam-combiners without loss, so that the other output intensity is simply $S_{21} = S_{12}(\phi_{im}(t) + \pi)$.

Developing the previous equation, we get

$$\begin{aligned} S_{12}(\vec{\theta}) &= \frac{1}{2}|\mathbf{A}_1(\vec{\theta})|^2 + \frac{1}{2}|\mathbf{A}_2(\vec{\theta})|^2 + |\mathbf{A}_1(\vec{\theta})||\mathbf{A}_2(\vec{\theta})|\cos(\phi_{im})\cos(\arg\mathbf{A}_1(\vec{\theta}) - \arg\mathbf{A}_2(\vec{\theta})) \\ &\quad + |\mathbf{A}_1(\vec{\theta})||\mathbf{A}_2(\vec{\theta})|\sin(\phi_{im})\sin(\arg\mathbf{A}_1(\vec{\theta}) - \arg\mathbf{A}_2(\vec{\theta})) \ . \end{aligned} \tag{3.12}$$

We immediately see that the first two terms are not modulated, and do not contribute to the synchronous detection. This inefficiency of internal modulation is one of its main limitations.

Distinction vs. exo-zodiacal signals

Let us assume a centro-symmetric exo-zodiacal light emission, that is the approximation of an homogeneous dust distribution seen with an arbitrary inclination i. The overall signal due to such an extended centro-symmetric source can be written as a sum of terms like $S_{12}(\vec{\theta}) + S_{12}(-\vec{\theta})$.

Now, *if the generalized entrance pupil of each sub-nulling interferometer is real*, equation (3.2) gives $\mathbf{A}_{12}(-\vec{\theta}) = \overline{\mathbf{A}_{12}}(\vec{\theta})$, so that $\arg\mathbf{A}_{12}(-\vec{\theta}) = -\arg\mathbf{A}_{12}(\vec{\theta})$. Since the sine is an odd function, the sine contribution in $S_{12}(\vec{\theta}) + S_{12}(-\vec{\theta})$ will cancel out, and the overall detected signal is then:

$$S_{12} \propto a + b\cos(\phi_{im}(t)) \ , \tag{3.13}$$

where a and b are real constants.

According to the last equations, the useful part of the modulated signal, specific to a planetary point-like source, is given by:

$$|\mathbf{A}_1(\vec{\theta})||\mathbf{A}_2(\vec{\theta})|\sin(\phi_{im})\sin(\arg\mathbf{A}_1(\vec{\theta}) - \arg\mathbf{A}_2(\vec{\theta})) \ . \tag{3.14}$$

This shows that, by setting $\phi_{im}(t)$ alternatively to $\pm\pi/2$, not only is the modulation efficiency (3.14) maximal for the planetary signal, but also the exo-zodiacal light not modulated at all! Since a 50:50 beam-combiner theoretically introduces an achromatic $\pi/2$ phase shift between the two input beams on one output, and $-\pi/2$ on the other (see section 1.2.3), the time-varying phase shift can be achieved by alternately detecting the two outputs of a single beam-combiner whose input beams are $\mathbf{A}_1$ and $\mathbf{A}_2$. In the following discussions, these two outputs will be referred to as S_{12} (with $\phi_{im} = \pi/2$) and S_{21} (with $\phi_{im} = -\pi/2$). With these notations, the definition for *modulation efficiency* reads:

$$\frac{|S_{12} - S_{21}|}{I_{\max}} = \frac{2\,|\mathbf{A}_1(\vec{\theta})|\,|\mathbf{A}_2(\vec{\theta})|}{I_{\max}}\,|\sin(\arg\mathbf{A}_1(\vec{\theta}) - \arg\mathbf{A}_2(\vec{\theta}))| \ , \tag{3.15}$$

where $I_{\max}$ represents the maximum detectable intensity coming from the planet (proportional to the overall collecting area). Note that this efficiency is zero if both $\mathbf{A}_1$ and $\mathbf{A}_2$ are real, or if any of the sub-nulling interferometers provides a null response at $\vec{\theta}$.

The two signals S_{12} and S_{21} can either be detected alternately on a unique detector, or continuously detected on two separate detectors. In the last case, two high frequency chopped signals can be constructed, depending on the alternation "$S_{12} - S_{21} - S_{12}$" or "$S_{21} - S_{12} - S_{21}$" (see figure 3.12). When demodulated, these two signals are completely equivalent. This redundancy improves the SNR by a factor $\sqrt{2}$, but it forces to use two detectors instead of a single one.

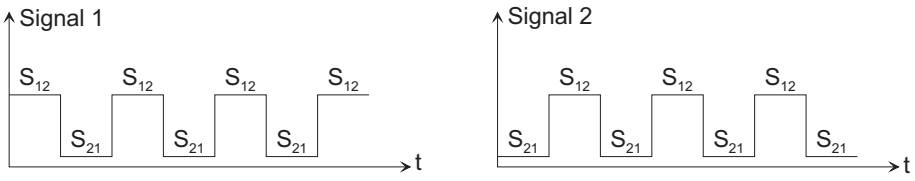


Figure 3.12: Two signals are constructed with outputs S_{12} and S_{21}. Only the useful planetary part of S_{12} and S_{21} is represented here. Background signals are much larger than the planet signal, but are not modulated at all since they have the same value in the two outputs. Note that the time variation of the original planet signal is much slower than the modulation period.

3.3 Current configurations with internal modulation

In this section, the principle of internal modulation is illustrated with some important examples.

3.3.1 The Mariotti nulling interferometer

The Mariotti configuration consists of six identical 1.5 m telescopes arranged on an equilateral triangle located at 1 AU from the Sun (at the L2 Sun-Earth point). It uses internal modulation between three DAC sub-arrays (figure 3.13). It has been studied in detail by Mennesson and Léger [39], following an original idea from Jean-Marie Mariotti.

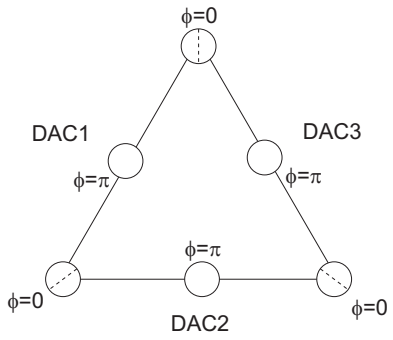


Figure 3.13: Geometry of the Mariotti nulling interferometer. Dotted lines indicates telescopes which are shared between adjacent sub-arrays. Note that the area ratio is 1:2:1 for each DAC, so that the diameter ratio is $1{:}\sqrt{2}{:}1$, in agreement with the remark on DAC's recombination (section 3.2.2). The recombination scheme of figure 3.8 is used for each individual DAC with equal size telescopes and 50:50 beam-splitters behind each summit telescope.

In that configuration, the telescopes of each DAC are recombined to produce a nulled output with θ^{-4} rejection. These nulled outputs are then recombined by pair, with a variable phase shift $\phi_{ij}(t)$ for the (i, j) pair.

This configuration has been further investigated by Alcatel Space [34]. It was found during the study that the 3-DAC concept suffered from the need for either long delay lines, or complex optical paths. These considerations led to the proposal of a modified 3-DAC, referred as to the "3D Mariotti" concept (see figure 3.14) in which an additional central hub is used for beam combination, and the free flyers are arranged in different planes. With that configuration, long delay lines are not required any longer, and the beam transfer optics is simplified. Unfortunately, the 3D configuration was found to suffer from thermal coupling due to the radiation from the sunshields of the upper-plane telescopes (see Thomas [54] p.92 for a complete discussion). That is the reason why *non plane configurations are not considered any longer.*

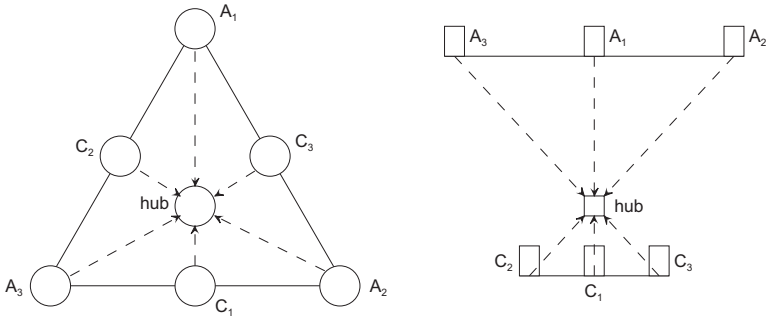


Figure 3.14: 3D Mariotti combination scheme. The six telescopes are lying on a paraboloid, the focus of which is occupied by the central hub.

3.3.2 The Robin Laurance nulling interferometers

As defined by Anders Karlsson [24], we shall refer to a Robin Laurance interferometer when the following conditions are satisfied:

- All telescopes are at equal distances from the beam-combiner spacecraft (no long delay line required),
- All telescopes are in one plane (perpendicular to the line-of-sight),
- Star light suppression is proportional to θ^4 (or better),
- All the telescopes are of the same diameter, but possibly seen with different transmission efficiencies at the recombination stage,
- Internal modulation is applied between at least two sub-nulling interferometers, with a $\pm\pi/2$ time variable phase shift,
- The only achromatic phase shift applied within a given sub-nulling interferometer is 0 or π.

The different solutions are referred to as RL$x(m_1,m_2, \ldots,m_n)$ where x is the number of sub-arrays (generally GACs), n is the number of telescopes and m_i the relative telescope size, starting with the largest telescope.

In the following paragraph, a complete study of RL3(3,2,0,1,0,2) is given. This hexagonal configuration has been designed by the ESA Darwin study manager Karlsson (see [24]) and further studied by Mennesson [38].

The RL3(3,2,0,1,0,2) configuration

It is the first configuration of the RL family. The good nulling properties of this configuration have rapidly been recognized, and this recombination scheme has been selected as a baseline for the current Darwin study. The RL3(3,2,0,1,0,2) has been fully studied by B. Mennesson [38]. The principal results are presented below, and will serve as a reference for the following discussions.

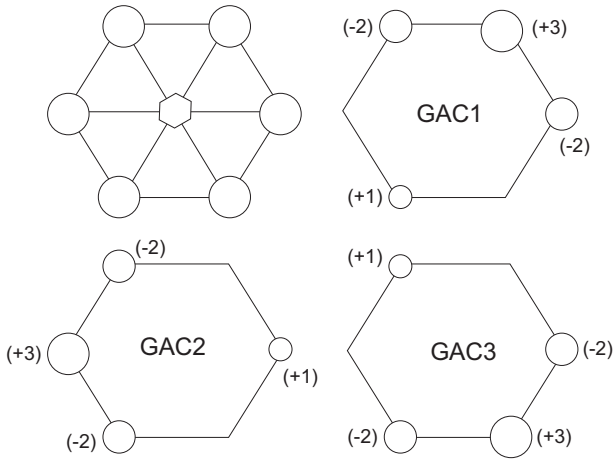


Figure 3.15: The RL3(3,2,0,1,0,2) interferometer and its three sub-nulling arrays (Generalized Angel Crosses) with their generalized entrance pupils.

Figure 3.15 shows the geometry of the array and the three sub-nulling GACs that compose it. Let L be the radius of the circle with the six telescopes, and let us define the reduced coordinates

$$x = \frac{2\pi L\theta \cos\phi}{\lambda} \quad , \quad y = \frac{2\pi L\theta \sin\phi}{\lambda} \, , \tag{3.16}$$

$$u = \frac{2\pi L\theta \cos(2\pi/3 - \phi)}{\lambda} \quad , \quad v = \frac{2\pi L\theta \cos(4\pi/3 - \phi)}{\lambda} \, . \tag{3.17}$$

Using the recombination scheme proposed by Karlsson and Mennesson, the intensities detected on the three GACs nulled outputs, resulting from the interference of four light

beams, are

$$I_{GAC1} = \frac{I_0}{36}| - 2e^{jx} + 3e^{-jv} - 2e^{ju} + e^{jv}|^2 \ , \tag{3.18}$$

$$I_{GAC2} = \frac{I_0}{36}| - 2e^{ju} + 3e^{-jx} - 2e^{jv} + e^{jx}|^2 \ , \tag{3.19}$$

$$I_{GAC3} = \frac{I_0}{36}| - 2e^{jv} + 3e^{-ju} - 2e^{jx} + e^{ju}|^2 \ , \tag{3.20}$$

where I_0 is the overall intensity seen by each telescope, including reflection loss and detector quantum efficiency. The recombination efficiency of any GAC is 8/9, corresponding to the maximum of $I_{GAC1}/2I_0$, where $2I_0$ is the maximum intensity theoretically detectable on one GAC output.

The transmission map of GAC1 is shown on figure 3.16. All plots are for a wavelength of 10 μm and a baseline of 30 m (i.e., the distance between any telescope and the beam-combiner is 15 m). The field of view is 500×500 mas (i.e., 500 mas in diameter). These are the same values as in [24] and [38], in order to allow simple comparisons. Note that the 30 m baseline is well suited to detect habitable planets, which are typically 80 mas away from a star located at 10 pc from Earth[11].

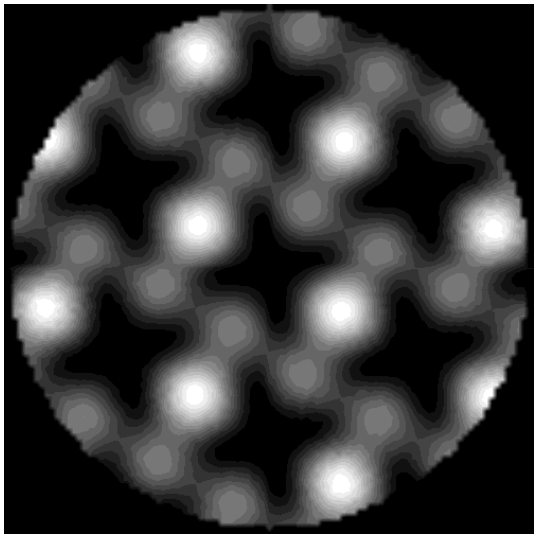


Figure 3.16: Transmission map of GAC1 on the sky. Field of view is 500×500 mas, interferometric array radius 15 m, and observing wavelength 10 μm.

When recombining the nulled outputs of GAC1 and GAC2 on a lossless beam-splitter, assuming $\phi_{im} = \pm\pi/2$ as explained in paragraph 3.2.5, we obtain the two complementary outputs $S_{12}(\phi_{im} = \pi/2)$ and $S_{21}(\phi_{im} = -\pi/2)$ shown on figure 3.17. The modulation

[11]The apparent projected distance between an Earth-like planet and its host star is actually unknown. The probability of finding a planet at a projected distance smaller than x times its true orbital radius is computed in [39]. Its is proven that 80% of the cases are between $x = 0.6$ and $x = 1$. Consequently, the typical angular separation of a Sun-Earth system at 10 pc is 80 mas.

efficiency (3.14) between GACs 1 and 2 is defined as the useful planetary part of the signal, and given by $|S_{12} - S_{21}|/2I_0$. The total modulation efficiency for the three GACs is finally defined as the mean of the modulation efficiencies for the three pairs of GACs:

$$\frac{|S_{12} - S_{21}| + |S_{13} - S_{31}| + |S_{23} - S_{32}|}{6I_0} . \tag{3.21}$$

That expression yields the following values:

- Mean modulation efficiency: 0.121, this is the relevant figure when the planetary location is unknown (detection phase).
- Maximum modulation efficiency: 0.299, this is the relevant figure once the planet has been detected and located. It is then possible to optimize the array size and orientation for spectroscopy at a given wavelength, with this efficiency.
- RMS of the modulation efficiency over the field of view: 0.102.

The modulation efficiency map is illustrated in figure 3.18. This map shows azimuthal directions of null modulation efficiency. At any wavelength, a source with that particular azimuth is undetectable, since its signal is not modulated at all. To overcome this problem, it is mandatory to rotate the array, which will make the sky coverage more uniform. Figure 3.18 shows the improvement of the sky coverage obtained by using for instance three different array positions, deduced from three 20° successive rotations. Thus, "azimuthal holes" have a bad influence since rotations are once again needed, but this is far less harmful than the continuous rotation needed when no internal modulation is used (see section 3.2.3). So, a secondary goal in the following study is to reduce the number and amplitude of rotations. Note that this kind of azimuthal holes are also present in the Mariotti configuration.

Other RL configurations

Three other RL configurations have already been studied: two of them are hexagonal and the last one has an irregular pentagonal shape. The most promising of them seems to be the RL2(0.70,1,0.84,0.54,0) pentagon. All those configurations will be described in the following chapter, and compared with the previous case.

3.3.3 Linear configurations

The TPF mission concept

Before describing the linear configurations, let us briefly present the Terrestrial Planet Finder (TPF) mission. This future NASA mission has exactly the same goals as Darwin, and uses the same technique (infrared nulling interferometry). So, there is a strong resemblance between these two missions, which could finally lead to a common international mission.

As for the Darwin mission, the first concept for TPF did not use internal modulation, and an orbit located at 5 AU from the Sun was selected in order to avoid thermal

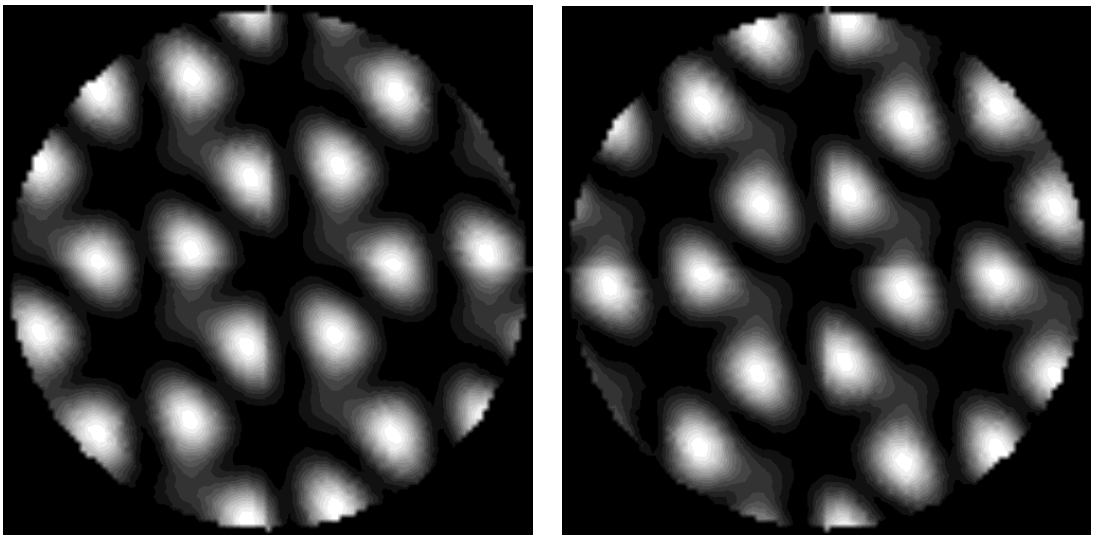


Figure 3.17: Transmission map of nulled output of GAC1 recombined with nulled output of GAC2, with a phase shift of $\pi/2$ (left) and $-\pi/2$ (right). Conversely to the transmission map of an individual GAC, these maps are asymmetric.

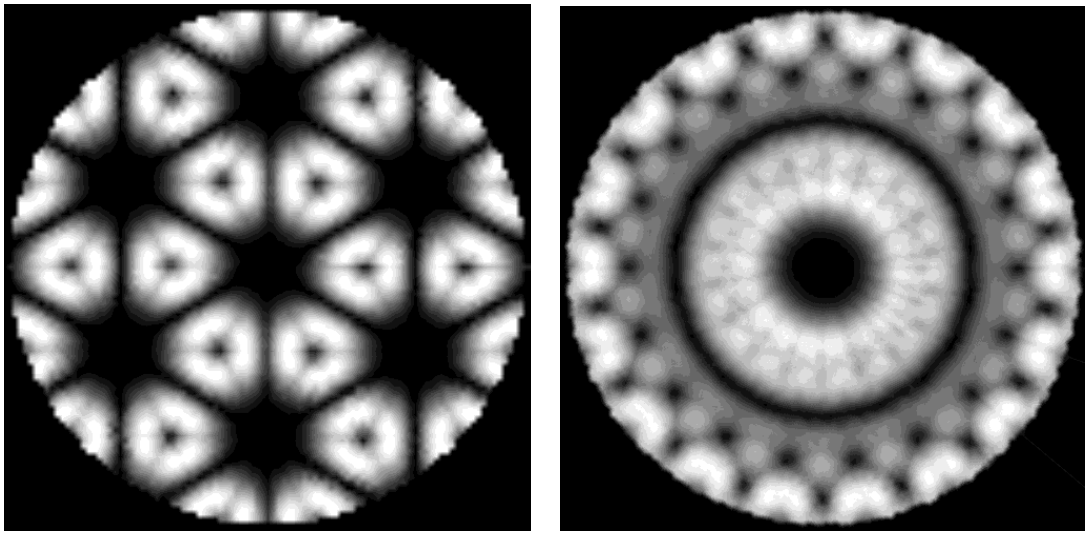


Figure 3.18: Left: modulation efficiency map. Right: mean modulation efficiency when superimposing three maps obtained by 20° successive rotations. Note that azimuthal directions of zero modulation are not present any more.

background from the zodiacal dust in our own solar system. The basic concept was a 4-telescope linear array dubbed OASES, yielding a θ^{-6} rejection rate (see appendix C.2.2). This configuration comes from the superposition of two Bracewell interferometers with different spacing, whose outputs are combined with a 180° phase shift (see figure 3.19). The rotation of the array around the line of sight is then used to modulate the flux from an off-axis point-like source.

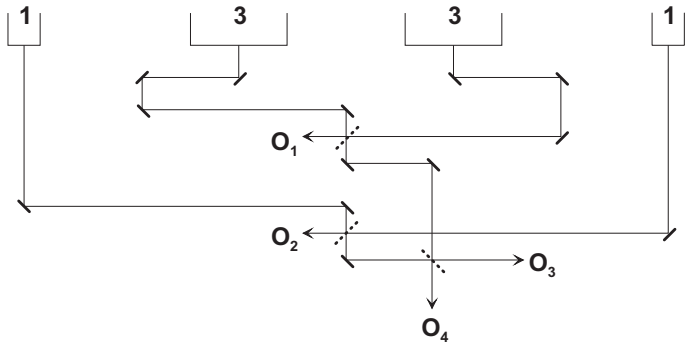


Figure 3.19: Schematic diagram of a 4-element linear nulling interferometer, adapted from [3]. The fixed beam-combining geometry is configured to give equal path lengths for all the four telescopes. The responses of the inner and outer pairs pass through zero on the median plane with equal and opposite slope, cancelling to high order (θ^{-6} rejection rate).

Evolution of the concept

The implementation of TPF has evolved in three ways (see [9]):

- Since large light-weight apertures are being developed for the NGST, it is now reasonable to consider using large apertures (about 3 m in diameter) for an interferometer operating in the higher background conditions at 1 AU.
- Using free-flying interferometer elements is now considered rather than a connected system of telescopes. It has the big advantage to enable the tuning of the baseline for each planetary system.
- In order to minimize the effects of long-term detector drift, and to reduce the need for rapid rotation of a free-flying constellation, more complex interferometer configurations using path length modulation and rapid chopping are being considered.

The first proposition for placing two nulling arrays close together, with their elements interdigitated is due to Woolf and Angel [59]. Their main goal was to eliminate the imaging ambiguity of 180° which tampers the usual linear configurations, and by the way, to cancel the exo-zodiacal emission. This can be done by combining the null outputs of the two sub-nulling arrays with an achromatic phase difference of 1/4 wave. The evolution of

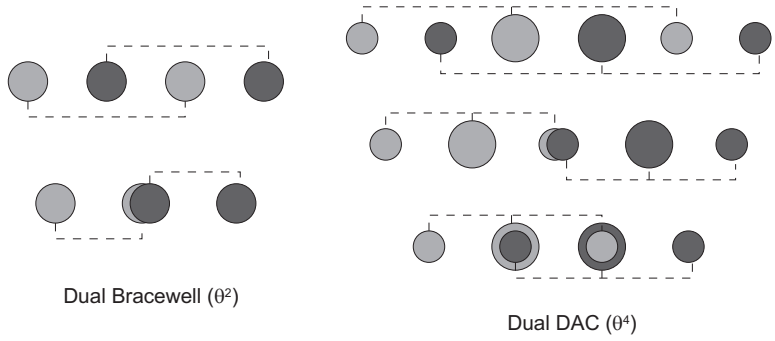


Figure 3.20: Dual linear nulling arrays (taken from [29]). Pairs of nulling interferometers can be combined in various ways to allow the rejection of unwanted noise sources.

this concept lead to a number of linear configurations with interlaced linear sub-arrays, summarized in figure 3.20.

The current baseline for TPF is a 4-telescope linear array called OASES, which could be used as a single nulling interferometer with a θ^{-6} rejection rate, or as a double DAC interferometer with internal modulation.

3.4 Feasibility of the Darwin mission

3.4.1 Observation strategy

Before proving the feasibility of the Darwin mission, an observation strategy must be chosen. The choice of the orbit is particularly important because it has a great influence on the noise level.

Choosing the orbit

We have seen in section 3.2.2 how the 1993 proposal of a 5 AU orbit has finally made the Darwin concept technically achievable (with 1.5 m-class telescopes instead of 8 m-class ones). But it was soon proven that such a distant orbit would be unrealistic for the Darwin mission because of insufficient payload mass capability and excessively long travel times.

Recently, the 1 AU orbit (at the Sun-Earth L2 point) has been re-evaluated, in the light of new elements:

- The advances in large cryogenic optics resulting from NGST enable the 1 AU mission to match the S/N of the previously studied 5 AU mission while simplifying many aspects of the mission design, particularly power systems.
- The actual brightness of the zodiacal cloud has been recalculated based on the latest dust models, and a decrease by a significant factor (about 4 at 10 μm) has been found.

- The 1 AU orbit decreases the orbital period, providing multiple opportunities to view each target over the lifetime of the mission, and drastically reduces the time required for Darwin to reach its operational orbit and begin taking science data.

These advances should allow the use of 1.5 m telescopes at a 1 AU orbit, as will be proven hereafter. Moreover, as stated by Woolf *et al.* [61], if other planetary systems have more dust emission from the 1 AU range than our solar system, then an interferometer at 5 AU would require greatly increased observing time, or increased telescope size. And indeed, if exo-zodiacal dust was much brighter than the solar system dust emission at 1 AU, mirrors of the same size would be needed whether Darwin (or TPF) was at 1 AU or 5 AU.

Observation sequence

A possible observation sequence is proposed in [39]. It consists in two successive phases:

- *Detection phase.* All spectral bins will be used simultaneously for detection. The global detection efficiency is therefore the mean value of the modulation efficiency averaged on azimuthal position and wavelength, weighted by source spectrum and wavelength dependence of quantum detector efficiency. Optimizing the interferometer baseline for detection of a potential Earth-like planet yields a modulation efficiency of about 12% when the RL3(3,2,0,1,0,2) configuration is used.

- *Spectroscopic phase.* Once the planetary location is known, the baseline and orientation of the array can be chosen to yield the maximum modulation efficiency (29.9%), at a selected wavelength, during spectroscopy.

3.4.2 Signal-to-noise calculations

The following calculations are largely based on a series of documents including the Darwin and TPF reports ([18] and [9]), papers from Mennesson and Léger [39], Beichman and Velusamy [8], and Thomas' thesis [54]. All the calculations are carried out for the RL3(3,2,0,1,0,2) configuration with an array radius of 15 m and a observation wavelength of 10 μm. A Sun-Earth system at 10 pc is assumed as a typical target. Note that in practice the radius of the interferometer has to be larger than 25 m in order to avoid thermal coupling between the satellites. A value of 15 m has been used throughout this work for reasons of easy comparison with previous results. This does not significantly change the conclusions of this work since observing an exo-system from 10 pc with an array radius of 15 m is completely equivalent to observing the same exo-system from 16.6 pc with an array radius of 25 m.

The planet signal

In order to calculate the planet signal, some scientific constraints must first be evaluated. The spectral range stretches from 6 to 18 μm in order to comprise the H_2O, CO_2 and O_3 spectral signatures (see figure 3.21). The spectral resolution R varies between 1 (detection phase) and 20 (O_3 spectroscopy at 10μm). The minimum signal-to-noise ratio (SNR)

allowing planet detection is about 5, while O_3 analysis needs a SNR of about 10. These values are taken from the DARWIN study report [18]. Our goal is here to show that the required SNR can be reached in a reasonable time in these two limiting cases.

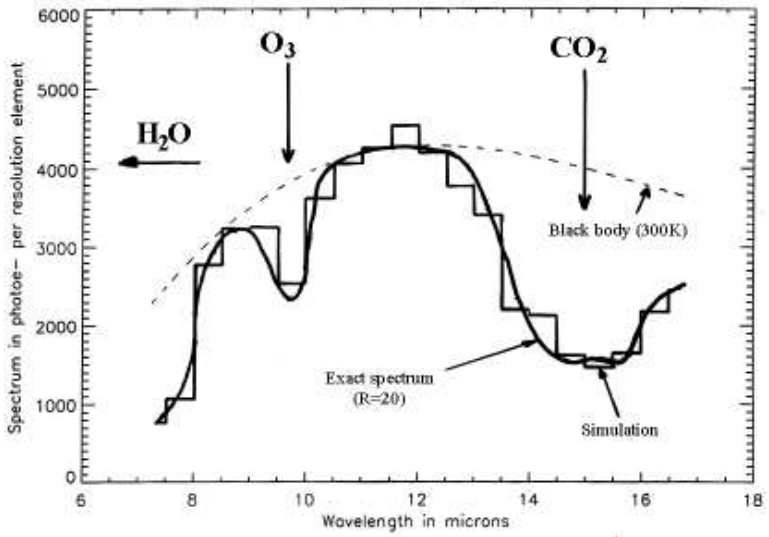


Figure 3.21: The Earth's spectrum as it would appear observed with DARWIN from a distance of 10 pc, and with a spectral resolution $R = 20$. This resolution allows the detection of the O_3 spectral signature.

Let us first compute the planet signal detected by DARWIN after synchronized demodulation of the interferometer outputs. Denoting ξ the modulation efficiency, the expression of the planet signal in photo-electrons per second (ph-e^-/s) reads

$$Q_p = 6\xi S \frac{\eta \Delta\nu}{h\nu} F_{\nu,p} \ , \tag{3.22}$$

where the factor 6 accounts for the six telescopes with surface S, $\eta \simeq 0.139$ is the instrumental efficiency[12], $\Delta\nu$ the bandwidth of the observation and $h\nu$ the quantum energy at observing frequency ν. The flux $F_{\nu,p}$ from an Earth-like planet at 10 pc is 0.34 μJy (see [8]). Writing this expression, it has been supposed that the output signals S_{ij} are all continuously detected on separate detectors, in order to obtain an optimum SNR ratio.

The noise contributors

As we have seen in section 3.2.5, the use of internal modulation at a frequency around 0.1 Hz reduces low frequency drifts of the background: the noise signals (physical background, detector response, stellar leaks, interferometer thermal emission) have a zero mean value but finite standard deviation after synchronized demodulation. In order to compute this

[12]The instrumental efficiency includes optical transmission Tr $\simeq$ 33%, detector quantum efficiency QE $\simeq$ 65% (SiAs detector) and field stop coupling efficiency 65%, so that the global efficiency is 0.139 (values from Karlsson, personal communication).

standard deviation, the amount of light from each background source coming through the null of an individual GAC must be computed. The noise at each GAC output is simply the standard deviation of the mean flux. The useful signal $S_{ij}(t)$ (from which the background flux is cancelled by demodulation) is obtained by combining the outputs of two GACs i and j with a time varying phase shift $\phi_{ij}(t)$. It can be easily proven that the noise is not affected by this coherent beam-combination: the level of noise associated to the detected signal $S_{ij}(t)$ is the same as the noise at each GAC output.

There are three types of noise contributors: the first type is associated to the interferometer's location (local zodiacal light, galactic cirrus), the second type to the observed star (stellar leaks, exo-zodiacal light) and the third type to the instruments (thermal emission of the interferometer, detector noise). Their contribution to the total noise is computed hereafter.

- *Local Zodiacal Cloud.* The contribution of the local zodiacal cloud is proportional to the field stop optical étendue $S\Omega$ (where S is the area of a telescope and Ω the solid angle of the field). The field stop was first designed for spatial filtering (see section 3.4.3), with an optical étendue $3.7\lambda^2$ defined by the first dark ring in the Airy pattern of the point spread function. Now, when the diaphragm radius θ is reduced from the first Airy zero, the flux from the Airy pattern first decreases slowly whereas the flux from the uniform zodiacal source decreases as θ^2. An optimum SNR is found for $S\Omega = 1.096\lambda^2$ (see [18]). If $R(\theta, \phi)$ represents the response of an individual GAC (see equation 3.4), the contribution in photo-electrons per second (ph-e$^-$/s) from the zodiacal cloud to an individual GAC output is[13]

$$Q_{\mathrm{LZ}} = \frac{\eta \Delta\nu}{h\nu} \int\int B_{\nu,\mathrm{LZ}} R(\theta,\phi)\theta \, \mathrm{d}\theta \, \mathrm{d}\phi \; . \tag{3.23}$$

  The latest dust models give a mean value of $B_{\nu,\mathrm{LZ}} = 6.19$ MJy/sr at 10 μm over the Darwin observation cone (see [18]).

- *Infrared Cirrus.* The galactic infrared cirrus is an additional background against which Darwin must observe. It has the same expression as the local zodiacal light. A surface brightness of 0.08 MJy/sr is assumed in the following calculations (see [54]).

- *Background Confusion Noise.* This contribution, due to star and galaxy infrared emission, will not be an issue at the high spatial resolution of Darwin (see [9]).

- *Exo-Zodiacal Emission.* In order to estimate the contribution of exo-zodiacal light to the photon noise, we use a model developed by Good (see Thomas' thesis [54]), based on the IRAS survey of the solar system zodiacal dust. In that model, the zodiacal cloud is considered as a grey body with emissivity $\varepsilon(r) = 1.114 \times 10^{-7} r^{-0.8}$, where r is the distance from the star in AU. The dust equilibrium temperature is given by $T(r) = 266\, r^{-0.36}$ K. Multiplying the black body brightness by the grey emissivity,

[13] Note that the number of telescopes and their collecting area are now included in the GAC response $R(\theta, \phi)$.

we obtain the exo-zodiacal cloud surface brightness:

$$B_\lambda(r) = \varepsilon(r)\frac{2hc^2}{\lambda^5}\frac{1}{\exp(\frac{hc}{\lambda kT(r)}) - 1} \ \ \mathrm{W/m^2/m/sr}\ . \tag{3.24}$$

The exo-zodiacal cloud contributes photon noise to the total signal starting at the dust sublimation radius and extending out to the edge of the primary beam of a single telescope. The dust destruction temperature is about 1500 K, so that the sublimation radius is 8.2×10^{-3} AU. Moreover, due to the interferometric transmission map, the hot inner part of the exo-zodiacal cloud is partially rejected by the interferometer (see figure 3.5). The total contribution of the exo-zodiacal cloud at a GAC output reads:

$$Q_{\mathrm{EZ}} = \frac{\eta\Delta\lambda}{h\nu}\int\int B_{\lambda,\mathrm{EZ}} R(\theta,\phi)\theta\,\mathrm{d}\theta\,\mathrm{d}\phi\ . \tag{3.25}$$

Note that little is still known about the prevalence of such dust in the target systems of the Darwin mission. But recently, several nearby main sequence stars were shown to have warm circumstellar matter at a level > 100 times the solar system value. Therefore, it appears necessary that the target systems will have to be observed in advance of the Darwin mission (see chapter 5).

- *Leakage Signal.* First of all, let us prove that the RL(3,2,0,1,0,2) interferometer, with a θ^{-4} rejection rate, allows to observe the planet despite the huge star-to-planet flux ratio (about 5×10^6). Assuming that the surface brightness of the star is constant over the whole stellar disk, the rejection rate can be easily evaluated by integrating the interferometer response over the stellar disk:

$$\rho = 2S\frac{\pi\theta_*^2}{\int\int_{\mathrm{star}} R(\theta,\phi)\theta\,\mathrm{d}\theta\,\mathrm{d}\phi}\ , \tag{3.26}$$

where $\pi\theta_*^2$ is the solid angle subtended by the star, and where $2S$ accounts for the total collecting area of a single GAC (a third of the overall collecting area). This rejection rate is about 8.9×10^8 for a Sun-like star at 10 pc, and an interferometer radius of 15 m as usual. On the other hand, a 30 m-long Bracewell interferometer, with a θ^{-2} rejection rate, would yield $\rho \simeq 3.5 \times 10^3$, which is obviously insufficient. Neglecting surface brightness variations over the stellar disk, the expression of the stellar leakage signal reads:

$$Q_{\mathrm{leak}} \simeq \frac{\eta\Delta\nu}{h\nu}B_{\nu,\mathrm{star}}\int\int R(\theta,\phi)\theta\,\mathrm{d}\theta\,\mathrm{d}\phi\ , \tag{3.27}$$

where the surface brightness $B_{\nu,\mathrm{star}}$ of a Sun-like star is about 10^{17} Jy/sr.

- *Leakage Jitter.* An additional noise source comes from wandering of the null across the star due to variations in pointing of the phase center of the interferometer and to OPD variations, which make the rejection rate a highly time dependent variable. Large

leakage fluctuations would be particularly harmful at the typical chopping frequency of 0.1 Hz. Numerical simulations have proven that the rejection rate remains above 10^5 for differential pointing errors below 2 mas rms and OPD fluctuations below 8 nm rms (see [54]). With these specifications, the leakage jitter noise should not be issue (Karlsson, personal communication). Other limitations to the rejection rate are briefly discussed in section 3.4.3. A rejection rate $\rho = 10^5$ is assumed in the following calculations. The contribution of the stellar leaks at a GAC output finally reads:

$$Q_{\rm leak} \simeq 2S\frac{\eta\Delta\nu}{h\nu}F_{\nu,\rm star}\rho^{-1} \ , \tag{3.28}$$

with $F_{\nu,\rm star} = \pi\theta_*^2 B_{\nu,\rm star} \simeq 1.6$ Jy at 10 μm.

- *Thermal Emission.* The equilibrium temperature of the interferometer is calculated so that the associated thermal emission remains below the local zodiacal light contribution over the whole spectral range. This can be achieved with a telescope temperature of 40 K. Note that this temperature can be reached with passive cooling if all the telescopes are in the same plane. The contribution of thermal emission is significant only at the longest wavelengths of the spectral range.

- *Detector Noise.* The integration time between two consecutive samplings of the internally modulated signal must be large enough so that the detector readout noise is significantly smaller than unavoidable noises. Using a readout noise of 10 e^- rms, Mennesson and Léger [39] conclude that a modulation frequency of 0.1 Hz can be used without noise penalty. This frequency is still much larger than the typical 10^{-4} Hz of rotating arrays. Another property of detectors is their dark current $i_{\rm dark}$, which is expected to be about 5 e^-/s (see [8]).

Signal-to-noise ratio and integration time

The total noise is computed in the table below for the two different observation phases. It is simply the standard deviation of the total counts, i.e., its square root since Poisson statistics is relevant for photon counts.

Signal and Noise Sources		
Signal (ph-e^-/s)	**Detection**	**O_3 Spectro**
Planet at 10 pc	1.09	0.135
Local Zodiacal Cloud	871.5	43.55
Infrared Cirrus	11.25	0.563
Exo-Zodiacal Cloud	152.4	7.617
Stellar Leakage	94.30	4.715
Dark Current	5	5
Total Counts per sec.	1135	61.58
Noise in ph-$e^-/s^{1/2}$	33.7	7.85

Now, the last thing to do is to deduce the observation time needed to achieve the required SNR (5 for detection and 10 for O_3 spectroscopy). A SNR of 5 is achieved in about 2.4×10^4 s for detection, i.e., about 6.7 hours of integration, whereas 3.4×10^5 s are needed for spectroscopy, i.e., about 94 hours (4 days). These values are relevant for the typical case of a Sun-Earth system at 10 pc. The evolution of integration time with respect to the exo-system distance can easily be computed: assuming that the SNR is limited by the noise associated with the local zodiacal cloud, it can be written as

$$SNR \propto \frac{F_p D^2 t}{\sqrt{\sigma_{\mathrm{LZ}}^2 t}} \tag{3.29}$$

where F_p is the planet flux, D is the telescope diameter and t the integration time. In this expression, only F_p depends on the exo-system distance d ($F_p \propto d^{-2}$). So, for a fixed SNR, the integration time is proportional to d^4, but also to D^{-4}. An increase in the telescope size would therefore be very beneficial.

3.4.3 Optical and mechanical constraints

Limitations to the rejection rate

We have proven in the previous section that a rejection rate of order $\rho = 10^5$ is sufficient to allow Earth-like planet detection. With that value, the leakage noise remains below the unavoidable noise associated with the local zodiacal cloud. This result has to be translated into technical specifications. This work has been done by M. Ollivier [43], who has identified three main limitations to the rejection rate. Their influence is computed for the case of a Bracewell interferometer for the sake of simplicity. A conservative value of $\rho = 10^6$ is chosen for this study.

- *Differential Phase Errors.* If the phase difference $\phi_1 - \phi_2$ between the two input beams is not exactly equal to π, the on-axis starlight extinction is not perfect. If we set $\phi_2 = \phi_1 + \pi + \epsilon_\phi$, the maximum value of the phase error ϵ_ϕ that achieves the minimum rejection rate $\rho_{\min}$ satisfies the following condition:

$$|\epsilon_\phi| \leq \frac{2}{\sqrt{\rho_{\min}}} , \tag{3.30}$$

  which gives $|\epsilon_\phi| \leq 2 \times 10^{-3}$ rad. This requires high precision optical delays lines, with accuracies of about 1 nm (see [28], chap. 16). Differential phase errors are also induced by differential "tip-tilt" errors between the two telescopes: a maximum error of $\delta\theta = 1.27 \times 10^{-3} \lambda / D$ can be tolerated if D is the telescope diameter (see [37]).

- *Relative Beam Intensities.* The on-axis cancellation of stellar radiation requires two input beams of equal intensity. In practice, the equality is not exactly satisfied: there is always a intensity error ϵ_I such that $I_2 = I_1(1 + \epsilon_I)$. The rejection ratio will be greater than $\rho_{\min}$ if

$$|\epsilon_I| \leq \frac{4}{\sqrt{\rho_{\min}}} . \tag{3.31}$$

In order to obtain a 10^6 rejection rate, the maximum allowed amplitude mismatch is 4×10^{-3}.

- *Polarization Errors.* Assuming, as Ollivier [43], that the recombination will be separately done on each of the s and p polarization states, the polarization plane of the two incoming beams must be equal to within a maximum error

$$|\epsilon_\theta| \leq \frac{1}{\sqrt[4]{\rho_{\min}}} , \tag{3.32}$$

which gives $|\epsilon_\theta| \leq 3 \times 10^{-2}$ in our case. On the other hand, if the two polarization states are recombined simultaneously, it is critical to avoid effects such as "differential birefringence": differing s-p phase delays between the two interferometer arms. The limit on the allowable differential birefringence is $4/\sqrt{\rho_{\min}}$ (see [28]).

Spatial and modal filtering

The most stringent constraint on wavefront quality is certainly the phase constraint. In the absence of atmosphere, the main phase fluctuations are due to imperfect mirror polishing. Equaling the two beams' phases everywhere on the pupil to within $2/\sqrt{\rho}$ imposes that the wavefront distortions are controlled to within $\lambda/(\pi\sqrt{\rho})$, i.e., to within $\lambda/3100$ for a 10^6 rejection rate (which is equivalent to $\lambda_{\text{vis}}/220$). This wavefront quality is definitely not achievable for 1.5 m large mirrors. Another approach is therefore necessary.

The high spatial frequency defects of the incoming wavefronts of the interferometer can be efficiently corrected by spatial filtering for instance using simple pinholes. Ollivier and Mariotti [44] have shown that by placing a pinhole at the focus of each telescope, with an optical radius of roughly $\lambda f/D$ (i.e., the size of an Airy spot), the effects of dust scattering or high frequency ripples of polishing residuals can almost be eliminated, since their contribution to the total PSF is diluted over a large region of the focal plane (due to a well-known property of Fourier transforms). Spatial filtering is very efficient for small phase defects, because their energy is spread on a large region of the focal plane, but not for low order defects (tilt, defocus, astigmatism) for which the previous constraints still apply.

To overcome this problem, the use of a single-mode waveguide has been proposed by Mennesson *et al.* [41]. When a single-mode waveguide is used, the incoming wavefront excites only one mode of the guide, whatever the shape of the entrance wavefront. The amplitude profile of this mode is fully determined by the physical properties of the guide. The shape of the initial wavefront will only affect the amount of energy coupled into the guide, so that phase effects are traded against amplitude effects. As a result, all phase defects of the wavefront besides residual piston (i.e., imperfect OPD control) are corrected. Qualitatively it is well known that fringe visibility is less affected by photometric unbalance than by phase defects (see section 1.3.1). Indeed, the effect of amplitude mismatch on the mean rejection rate is of second order as compared to phase aberrations (see [41]).

The new constraints obtained in the case of modal filtering are far more realistic, and considerably improve the mission feasibility. In order to get a 10^6 rejection rate, the optical surface quality drops from $\lambda_{\text{vis}}/220$ rms to $\lambda_{\text{vis}}/3$ rms, and the pointing accuracy

from 1.2 mas rms to 38 mas rms. Note that, if amplitude mismatching was corrected in real time[14], then the expected rejection rate would only be limited by polarization mismatches in the optical train. Polarization effects could therefore well set the ultimate limit to the rejection rate. A detailed analysis of imperfections can be found in Wallner *et al.* [56], and symmetry requirements in Serabyn [51].

3.4.4 Technological challenges

Two main items require technological developments: interferometric nulling and control of spacecraft flying in formation. Both are under development for ground observation or precursor missions, but will also require some specific development. In this section, some technological challenges related to interferometric nulling are briefly discussed (based on the Alcatel IRSI-Darwin study [34]). Formation flying will not be described here (see [30] for more information on this topic).

Optical components

- *Dichroic mirrors, beam-splitters and compensating plates.* Precise manufacturing control of these components is essential in order to achieve the required uniformity of transmission and phase delay for both polarizations, over the full spectral bandwidth and at cryogenic temperatures.

- *Mid-IR integrated optics and monomode optical fibers.* Suitable monomode waveguide, capable of covering the full spectral bandwidth of Darwin (6 - 18 μm) with acceptable losses do not yet exist. An overview of technological possibilities is given in [37] (chap. 7). Modal filtering could be achieved with "long holes", hollow guides, integrated optics components or optical fibers. Monomode fibers are already used till 5 μm with attenuation below the dB/m level.

- *Polarizing components.* Thin polarizing and birefringent elements will be required with high transmission efficiencies across the whole spectral range.

Achromatic phase shifter

As we have seen, depending on the use of internal or external modulation, phase shifts equal to π rad or to a fraction of π are required. In any case, the phase shift to perform must be achromatic so as to preserve the full efficiency of the interference process over the whole spectral bandwidth. If this is not the case, the destructive interference is not complete and some light from the parent star is not suppressed, which could prevent the planet light from being detectable. Therefore an approach based on simply inserting an extra optical path between waves is not relevant, and dedicated systems must be used. Let us briefly present some achromatic phase shifting devices currently under study or test, which could achieve the required 2×10^{-3} rad phase precision over the whole bandwidth.

[14]This can be achieved by attenuating one of the beams by a polarizer, or by slightly tilting one of the wavefronts at the entrance of the waveguide.

- *Dispersive materials.* The wavelength dependence of a half wave optical delay can be matched by adding optical path length through a dielectric medium, such as a glass plate. This technique has been described and tested by Morgan and Burge [42]. The resulting phase difference as a balanced linear component and a second order curvature due to the wavelength dependent index of refraction of the dielectric material. A second dielectric material can be used to balance the second order curvature, yielding residual phase errors with third order curvature. The phase error could be further reduced by adding additional plates, but two dielectric plates seem to be sufficient. Mid-infrared observations of circumstellar environments have been carried out by Hinz *et al.* [21] using that technique with two co-mounted telescopes of the Multiple Mirror Telescope.

- *Birefringence.* This approach uses a set of plane parallel plates made of birefringent crystal and cascaded along the propagation axis, and oriented perpendicular to it. The principle is to perform a sequence of polarization changes, each providing a chromatic phase shift but each being different from the other (see [48]).

- *Focus crossing.* A beam crossing a focus is achromatically shifted by π, with respect to a parallel beam following an equal optical path. Nulling can be achieved with a two-aperture interferometer using this dephasing property in one arm, by means of a cat's eye device. This technique has been validated with the "Achromatic Interfero Coronagraph" (see section 2.3.4). Note that splitting the focus in two sub-foci (by means of cylindrical mirrors for example) can be used to achieve a $\pi/2$ achromatic phase shift (see [48]).

- *Field flip.* The use of two orthogonal rooftop mirrors in one arm of a Michelson interferometer can be used to create a π achromatic phase shift between the two interfering waves. A "rotational shearing interferometer" based on this technique has been devised by Serabyn [52]. The advantages of this technique include a reliance solely on flat mirrors, near-perfect symmetry with respect to the two polarizations, and insensitivity to the beam-splitter's exact reflection/transmission ratio (because it is used in double pass). The best null depth obtained with a laser-diode source is 8×10^{-6}, while a transient null depth of 2×10^{-4} has been achieved on a filtered white-light source of 8% bandwidth.

- *Total reflection.* This new approach is based on the dephasing properties of total reflection, and uses, on each interferometer arm, a couple of isosceles prisms having unequal refractive indexes (see [48] for details).

Other optical sub-systems

Some other optical sub-systems will require technology development efforts in the coming year in order to raise their performance to a level compatible with the mission requirements (see [34]). These include:

- the fringe sensor, which should be capable of sensing variations of external OPD with an accuracy of 0.75 nm at a frequency of 10 Hz,

- the flux matching device, whose performance has still to be evaluated,
- the delay lines, which are likely to require considerable development work before they can achieve the expected performance and reliability levels on a remote space-borne system operating at 40 K,
- the laser metrology system, which is required to stabilize the drift in differential OPD between pairs of telescopes and the beam-combining hub, and to enable the fringe sensor to detect fringes in the vicinity of the interferometric null condition.

Chapter 4

Further exploration of the DARWIN aperture configurations

In this chapter, a systematic study of the DARWIN aperture configurations is undertaken. A special interest is given to configurations with internal modulation, and particularly to RL configurations. The maximum number of telescopes will be limited to six, for obvious reasons of cost and complexity, and because of the limited room in an Ariane 5 fairing.

4.1 RL configurations with five telescopes

Before studying RL configurations[1] with five telescopes in detail, let us prove that it is the minimum number of telescopes required to perform internal modulation, if the telescopes are restricted to stand on a circle (at equal distances from the beam combiner). This fact has already been proven by Mennesson [38].

- Three telescopes: as proven in appendix C.1, there is no configuration with three telescopes on a circle allowing a θ^4 null. So, RL configurations with three telescopes cannot be constructed.
- Four telescopes: as proven in appendix B.2.2, for any given distinct positioning of the four telescopes, there is one, and only one, set of amplitude coefficients ensuring a θ^4 null at the recombination stage. So, there cannot be any internal modulation with four telescopes since two subsets of four telescopes are needed.

4.1.1 Modulation between two GACs

Let us first consider two sets of four telescopes (GACs), three of them being common to the two sets. An infinity of solutions exists since for each telescope position (i.e., for each set of polar angles δ_k), we can find a couple of GACs with three common telescopes.

This case has already been fully studied by Mennesson [38]. The calculations and results are summarized below, because they lead to important conclusions. All plots in

[1] see section 3.3.2 for the definition of an RL configuration

this chapter are for a wavelength of 10 μm and a baseline of 30 m (i.e., the distance between any telescope and the beam-combiner is 15 m). The field of view is 500×500 mas (i.e., 500 mas in diameter). These are the same values as in [24] and in chapter 3, in order to allow simple comparisons.

Equal size telescopes

This case is of practical importance, since the manufacturing of five telescopes with equal size will significantly reduce the costs. Mennesson has proven in his report [38] that one and only one set of amplitudes allows telescopes of equal sizes. Denoting α_k the amplitude associated with telescope k in GAC1, and α'_k the amplitude associated with telescope k in GAC2, this (normalized) set of amplitudes writes:

$$\begin{aligned}&\alpha_1 = 1/\sqrt{2} = 0.707\,,\ \alpha_2 = -1\,,\ \alpha_3 = k = 0.838\,,\ \alpha_4 = -\sqrt{1-k^2} = -0.546\,,\ \alpha_5 = 0\,,\\ &\alpha'_1 = 0.707\,,\ \alpha'_2 = 0\,,\ \alpha'_3 = -0.546\,,\ \alpha'_4 = 0.838\,,\ \alpha'_5 = -1\,.\end{aligned} \tag{4.1}$$

These are exactly the amplitude coefficients of the RL2(1,0.84,0.54,0,0.70) configuration, proposed by Karlsson and Mennesson in [24]. This configuration is a non-regular pentagon configuration, with telescope locations

$$\delta_1 = 0\,,\ \delta_2 = 1.069\,,\ \delta_3 = 2.455\,,\ \delta_4 = 2\pi - \delta_3\,,\ \delta_5 = 2\pi - \delta_4\,. \tag{4.2}$$

This is the one and only set of δ_k providing the right amplitude coefficients for equal size telescopes. This can easily be shown by numerically solving the nulling equations (3.8) for the two GACs, introducing the amplitude coefficients (4.1), with a Newton-Raphson method: the iterations always converge towards that particular set of δ_k or towards a non-physical solution with all telescopes at the same place.

Let us now describe the RL2(1,0.84,0.54,0,0.70) in more detail. The geometry of the pentagon array is shown in figure 4.1, with the generalized entrance pupil of the two GACs. A schematic recombination scheme for the beams is presented in figure 4.2. Let A_0 be the overall amplitude modulus seen by each telescope, and $I_0 = A_0^2$ the corresponding intensity, including reflection loss and detector quantum efficiency. From the recombining scheme, we get for GAC1 an amplitude modulus for each telescope given by:

$$\begin{aligned}\text{Telescope 1} &: A_0 \cdot \sqrt{1/2} \cdot \sqrt{1/2} \cdot \sqrt{1/2}\,,\\ \text{Telescope 2} &: A_0 \cdot \sqrt{1/2} \cdot \sqrt{1/2}\,,\\ \text{Telescope 3} &: A_0 \cdot k \cdot \sqrt{1/2} \cdot \sqrt{1/2}\,,\\ \text{Telescope 4} &: A_0 \cdot \sqrt{1-k^2} \cdot \sqrt{1/2} \cdot \sqrt{1/2}\,.\end{aligned}$$

Setting $t = 2\pi L\theta/\lambda$, the output amplitudes of the GACs read:

$$\mathbf{A}_{GAC1}(\theta,\phi) = \frac{A_0}{2}\left(\frac{1}{\sqrt{2}}e^{jt\cos\phi} - e^{jt\cos(\delta_2-\phi)} + ke^{jt\cos(\delta_3-\phi)} - \sqrt{1-k^2}\,e^{jt\cos(\delta_4-\phi)}\right)\,,$$

$$\mathbf{A}_{GAC2}(\theta,\phi) = \frac{A_0}{2}\left(\frac{1}{\sqrt{2}}e^{jt\cos\phi} - \sqrt{1-k^2}\,e^{jt\cos(\delta_3-\phi)} + ke^{jt\cos(\delta_4-\phi)} - e^{jt\cos(\delta_5-\phi)}\right)\,,$$

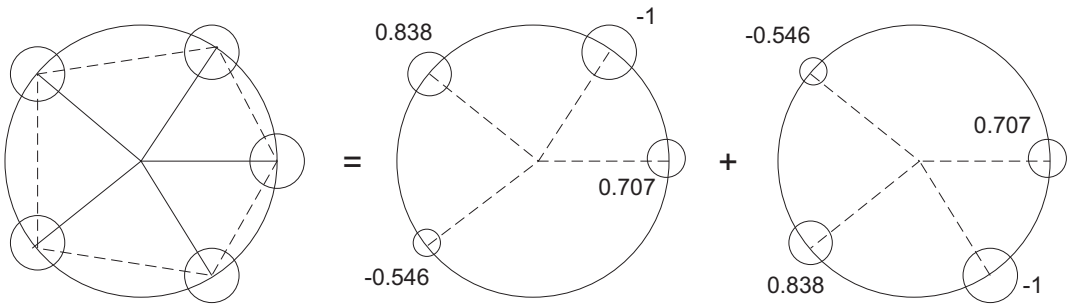


Figure 4.1: The RL2(1,0.84,0.54,0,0.70) nulling interferometer and its two sub-nulling interferometers, the GACs, with their generalized entrance pupils.

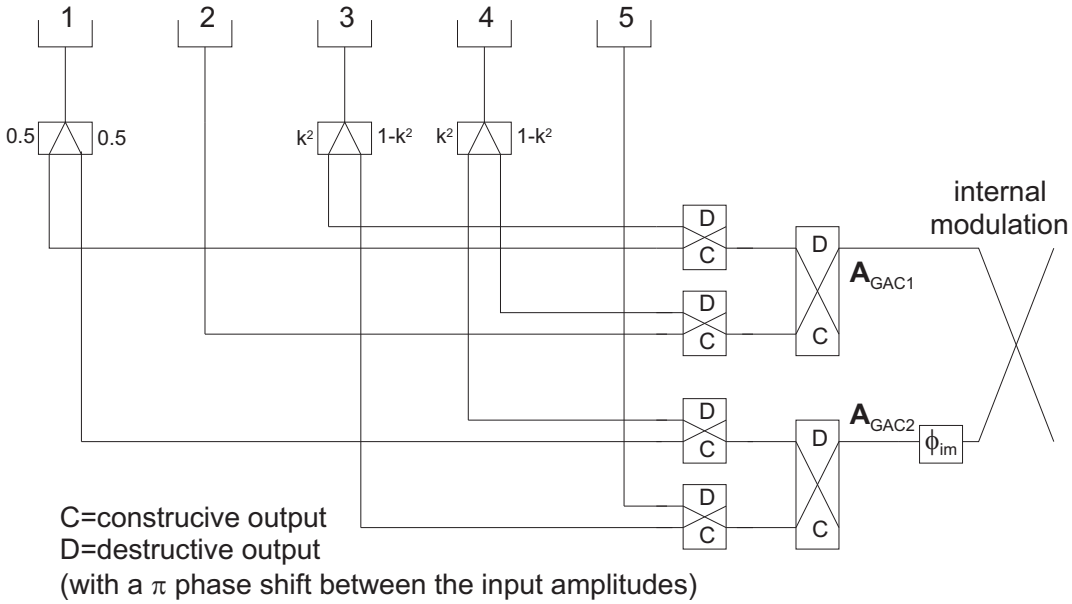


Figure 4.2: Schematic recombination of the beams for the RL2(1,0.84,0.54,0,0.70) interferometer. Note the beam-splitter intensity coefficients of k^2, $1-k^2$ (where $k = 0.838$).

where the $A_0/2$ factor accounts for the losses in the recombining scheme: one half of the intensity is lost at each beam-combiner since only one of the two outputs is used. The intensities are the square of the amplitude moduli. The normalized[2] intensities represent the transmission map of each GAC, plotted in figure 4.3. The recombination efficiency of any GAC, defined as the maximum of $I_{GAC}/2.5I_0$, is equal to 0.955.

The recombination of the nulled outputs of GAC1 and GAC2 on a lossless beam-combiner yields the two detected signals S_{12} and S_{21}, defined by equation (3.11). Assuming $\phi_{im} = \pm\pi/2$ as usual, figure 4.4 shows the transmission maps for the two cases.

The expected modulation efficiency as a function of the source location is given by:

$$\frac{|S_{12} - S_{21}|}{5I_0} \ . \tag{4.3}$$

The modulation characteristics are the following:

- Mean modulation efficiency: 0.118. This is the relevant figure when the planetary location is unknown (detection phase).
- Maximum modulation efficiency: 0.546. This is the relevant figure when the planet has been detected and located. It is then possible to optimize the array size and orientation for spectroscopy at a given wavelength.
- RMS of the modulation efficiency over the field of view: 0.117.

Figure 4.5 shows how the sky coverage can be improved by using for instance four different array positions deduced from 18° rotations. The modulation crown surrounding the central null becomes quite uniform with this trick, with a relative efficiency above 50% of maximum over a 40 mas-wide crown[3]. Three rotations would be enough to get rid of the azimuthal directions of zero modulation, but some regions would still have a quite low modulation efficiency (only about 20% of maximum). A compromise must be found between the number of rotations (i.e., the time and energy consumption in the detection phase) and the uniformity of the map (i.e., the quality of the SNR over the modulation crown).

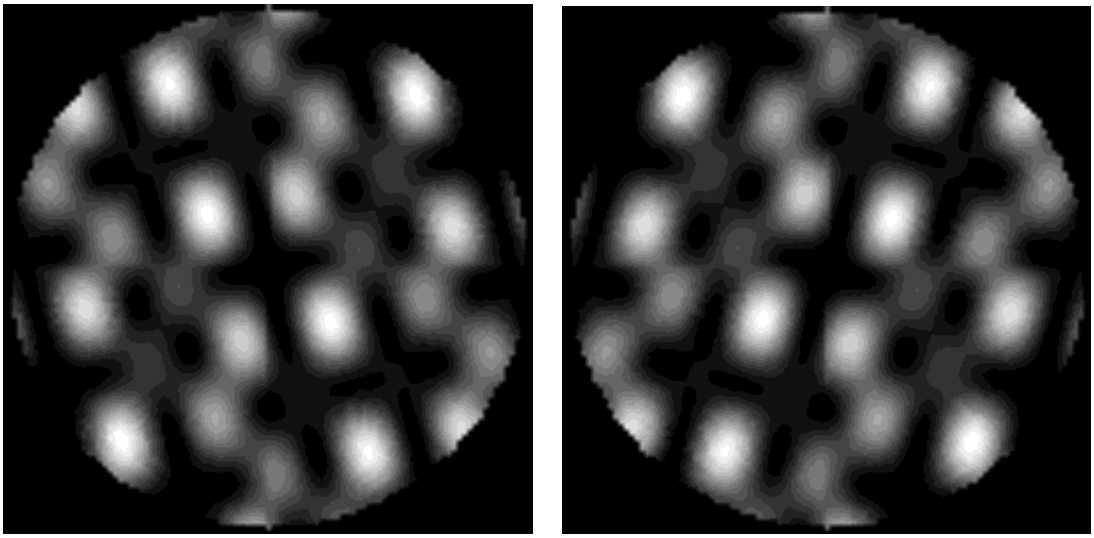
Different size telescopes

When it is not necessary to have similar telescope size, any set of angles δ_k can yield a nulling interferometer with internal modulation between two GACs. Such configurations are not members of the RL family, since the telescope diameters must be equal in RL configurations.

Mennesson has developed an IDL code to search for the optimal configuration of that kind (optimal means higher modulation characteristics in our case). The results of this optimization prove that no substantial gain is obtained by using telescopes of different sizes. Only a small gain in the mean modulation efficiency has been observed (from ~ 0.12 to ~ 0.14).

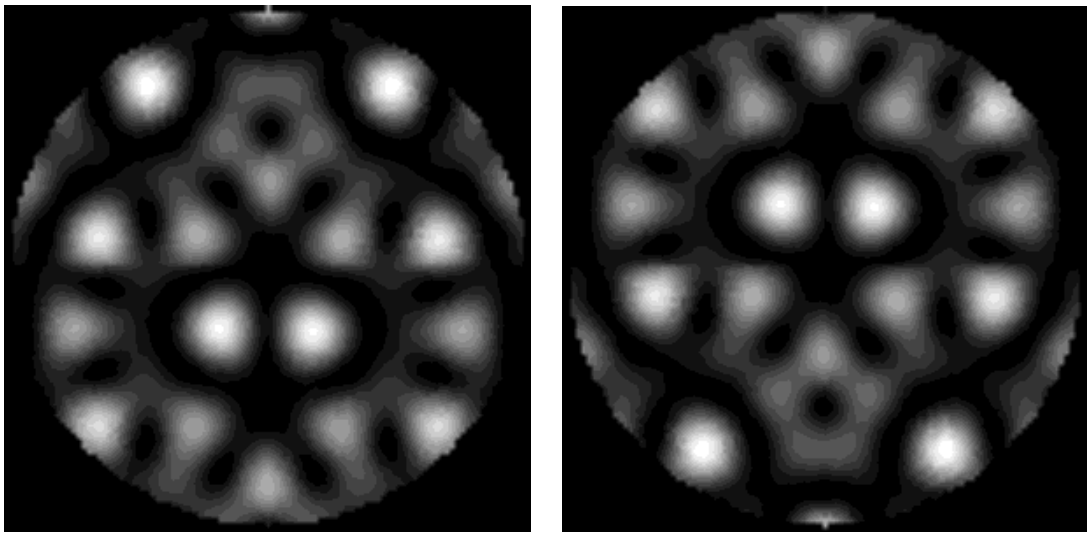
[2]Normalized means divided by $2.5I_0$, the maximum intensity detectable on one GAC output.

[3]If the star is located at 10 pc from the observer, a 40 mas-wide crown centered at 80 mas from the star typically represents a 0.5 AU-wide region at 1 AU from the star (for a random inclination of the system). This region covers a great part of the habitable zone for a G-type star. Thus the array is well suited for Earth-like planet detection.



Transmission map of GAC1 Transmission map of GAC2

Figure 4.3: RL2(1,0.84,0.54,0,0.70) configuration: transmission maps of the two GACs projected on the sky. Field of view is 500×500 mas, interferometric array radius 15 m, and observing wavelength 10μm.



Transmission map of nulled output S_{12} Transmission map of nulled output S_{21}

Figure 4.4: RL2(1,0.84,0.54,0,0.70) configuration: transmission maps of nulled output from GAC1 recombined with nulled output from GAC2, with a phase shift of $\pi/2$ (left) and $-\pi/2$ (right).

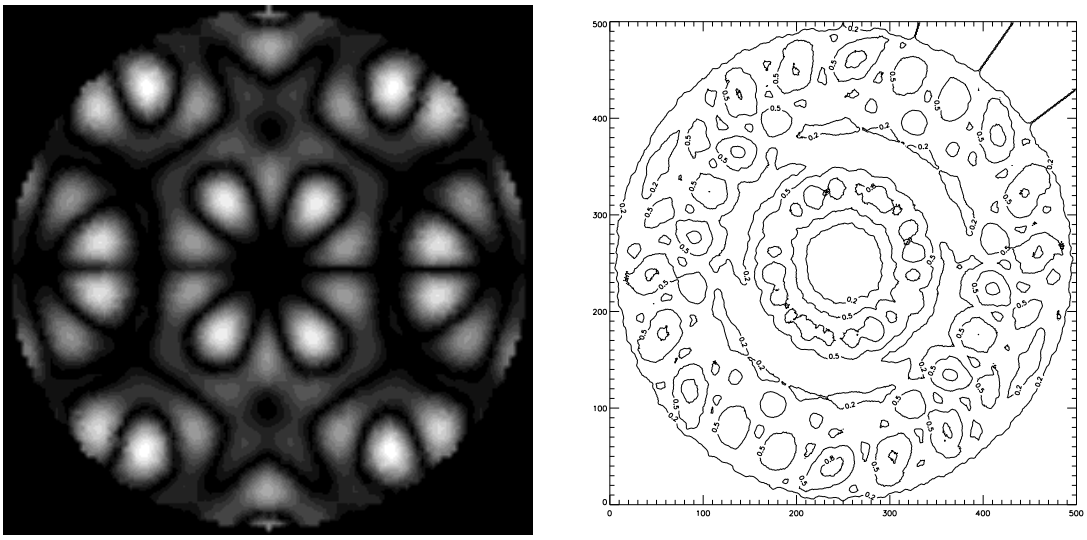


Figure 4.5: Modulation efficiency over the 500×500 mas field of view for the RL2(1,0.84,0.54,0,0.70) configuration. The right-hand map is deduced by superimposing four maps obtained by successive rotations of 18°. Isophotes at 20, 50 and 80% of maximum are shown.

Conclusion

The small increase in the modulation characteristics does not justify the use of different size telescopes. In fact, RL2(1,0.84,0.54,0,0.70) seems to be a very attractive configuration, since it uses the minimum number of telescopes (as long as linear configurations are not considered) and has a good modulation efficiency (compare to RL3(3,2,0,1,0,2) in section 3.3.2). This configuration is referenced as the "Karlsson" pentagon in the following discussions.

4.1.2 Modulation between two FTRCs

It is proven in appendix B.2.3 that when the positions of five telescopes are fixed on a circle, an infinity of real-pupil configurations using the five telescopes (each configuration having different amplitudes for the telescopes) produce a θ^4 on-axis extinction. In the following discussion, such configurations are called FTRCs (Five Telescopes Real Circle).

In this case, the two sub-nulling interferometers will both use the same five telescopes, so that the total number of telescopes is still five. In fact, this paragraph is a generalization of the previous one, since the GACs are only a particular case of FTRCs, with one of the telescope amplitudes equal to zero.

The identification of such configurations is a tedious task if one does not impose some additional constraints, because of the many degrees of freedom: the overall telescope sizes, the generalized pupil for each individual FTRC (i.e., the amplitude with which the telescopes are seen in each FTRC), and the positions of the telescopes on the circle. So, we will restrict the study to RL configurations, with all telescopes having the same overall

diameter. As the previous study showed that allowing different size telescopes does not significantly improve the modulation efficiency, this restriction should not be detrimental.

Bases of the study

As stated above, each set of five angular positions δ_k could lead to an RL configuration with five telescopes of same diameter placed on a circle. The first thing we have to do is therefore to choose some arbitrary values for the δ_k. The next step is to choose a first FTRC configuration based on that geometrical configuration. Among the infinity of other FTRCs with that geometry, we then search a FTRC which, superimposed on the first, would give equal diameters for the telescopes. Depending on the chosen set of δ_k, such a configuration will exist or not. In fact, the study shows that only some particular values for the δ_k lead to a solution. But when a solution exists, there seems to be an infinity of other solutions with the same geometry. These properties of solutions come from the complicated structure of a nonlinear system of 10 equations with 9 unknown parameters [4].

Recombination in detail

An important factor for the modulation characteristics of a given configuration is the recombination scheme for the sub-nulling interferometers, and the associated recombination efficiency. We have already discussed the recombination schemes for DACs and GACs (see section 3.2.2). A more exhaustive study of beam-combination is given in [40], which deals with co-axial[5] beam-combination in general.

It has already been shown[6] that, starting with the simple case of two beams recombined with a single beam-splitter, it is easy to deduce a scheme for four beams, and then, by recurrence, for 2^m beams. All beam-splitters are 50:50, and only one path connects one input to one output.

In the case of a system with five telescopes, since $5 = 2^2 + 1$, we need to consider a beam combiner of the third order ($m = 3$). An example of such a combiner is given in [40]. It uses eight inputs (three of which are "virtual"), and assumes that broad-band phase retardation devices are introduced in each beam of the interferometer. Each beam undergoes three reflections or transmissions, in order to equalize to moduli of the on-axis complex amplitudes (without it, on-axis light cancellation would not be possible). Assuming lossless beam-splitters, the reflection and transmission coefficients have both the same modulus $|r| = |t| = 1/2$, so that only 1/8th of each beam is actually used (one half of the intensity is lost at each beam-splitter). The efficiency of the proposed scheme is therefore only 5/8, while it approaches unity for a GAC, for which second order beam-combination is used.

[4]The ten equations are the six linear conditions for a θ^4 central extinction for the two sub-arrays and the four nonlinear conditions to get equal size telescopes. The nine unknown parameters are the nine beam amplitudes that remain free when an arbitrary amplitude is fixed.

[5]This type of beam-combination is always used in this work. It uses beam-splitter type arrangement, for which the beams are perfectly coincident, to produce an infinite interfringe distance (fluffed-out fringes).

[6]for example in some of the ESO studies for the VLTI

This recombination scheme will be used in the following discussions. Its poor recombination efficiency will be an intrinsic limitation to the modulation efficiency of the five-telescope configurations. A better scheme, with less losses, would be a great improvement not only for the five-telescope configurations, but also for the others, that also suffer from the losses at each beam-splitter.

Qualitative results

In order to search for configurations with equal size telescopes, a tolerance was fixed to defined what "equal sizes" means: in the following, a relative difference of 1% between the telescopes surfaces is tolerated. Our goal is to find some sets of δ_k which allow for equal size telescopes, and to study the modulation characteristics of those configurations.

As a result of a random search, about ten suitable δ_k sets were found. For each set, there are many possible amplitudes (i.e., beam-splitter ratios) for the two FTRCs, depending on the choice for the first FTRC. The modulation efficiencies of all those configurations were computed, and a somewhat strange property was found: for a given geometrical configuration with five telescopes on a circle, the modulation characteristics are exactly the same whatever the relative amplitudes in the recombination scheme, i.e., whatever the two chosen individual FTRCs.

For example, using the same angular positions as for the already studied RL2(1,0.84, 0.54,0,0.70) configuration, the IDL code gave more than one hundred solutions, even if the set of tested amplitudes was not very large (about 10000 tested configurations). Among those solutions, the RL2(1,0.84,0.54,0,0.70) was naturally present, since the GACs are special cases of FTRCs, with one diameter equal to zero.

Another special configuration is the regular pentagon (equal angular separation between successive telescopes). Detailed results for this configuration are presented in the following paragraph.

The regular pentagon

The angular positions δ_k of the telescopes are $\frac{2}{5}(k-1)\pi$ for a regular pentagon configuration. This set is one of the suitable sets of δ_k for a RL2 configuration to exist. Let us write three possible sets of amplitudes, among the infinity of possibilities:

$$\text{Set 1:} \quad \begin{cases} \text{FTRC1: } \alpha_1 = 1,\ \alpha_2 = 1.482,\ \alpha_3 = -3.398,\ \alpha_4 = 4.016,\ \alpha_5 = -3.1, \\ \text{FTRC2: } \alpha'_1 = 3.9,\ \alpha'_2 = -3.741,\ \alpha'_3 = 2.154,\ \alpha'_4 = 0.257,\ \alpha'_5 = -2.569, \end{cases}$$

$$\text{Set 2:} \quad \begin{cases} \text{FTRC1: } \alpha_1 = 1,\ \alpha_2 = -2.218,\ \alpha_3 = 2.589,\ \alpha_4 = -1.971,\ \alpha_5 = 0.6, \\ \text{FTRC2: } \alpha'_1 = -2.4,\ \alpha'_2 = 1.353,\ \alpha'_3 = 0.210,\ \alpha'_4 = -1.693,\ \alpha'_5 = 2.530, \end{cases}$$

$$\text{Set 3:} \quad \begin{cases} \text{FTRC1: } \alpha_1 = 1,\ \alpha_2 = -3.218,\ \alpha_3 = 4.207,\ \alpha_4 = -3.589,\ \alpha_5 = 1.6, \\ \text{FTRC2: } \alpha'_1 = -4.1,\ \alpha'_2 = 2.729,\ \alpha'_3 = -0.315,\ \alpha'_4 = -2.219,\ \alpha'_5 = 3.905\ . \end{cases}$$

The second of these configurations is illustrated in figure 4.6.

As explained above, these configurations give exactly the same modulation characteristics:

- Mean modulation efficiency: 0.060

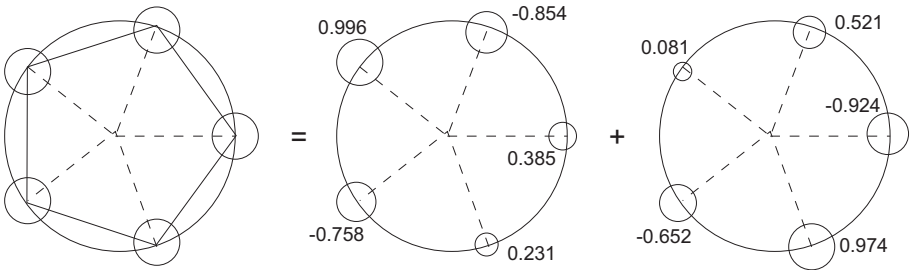


Figure 4.6: The regular pentagon nulling interferometer and a possibility for its two sub-nulling interferometers (FTRCs), with their generalized entrance pupils.

- Maximum modulation efficiency: 0.298
- RMS of modulation efficiency: 0.061

As expected, the modulation efficiency is about two times smaller than the modulation efficiency of the Karlsson pentagon (with internal modulation between two GACs), because of the recombining scheme is of the third order.

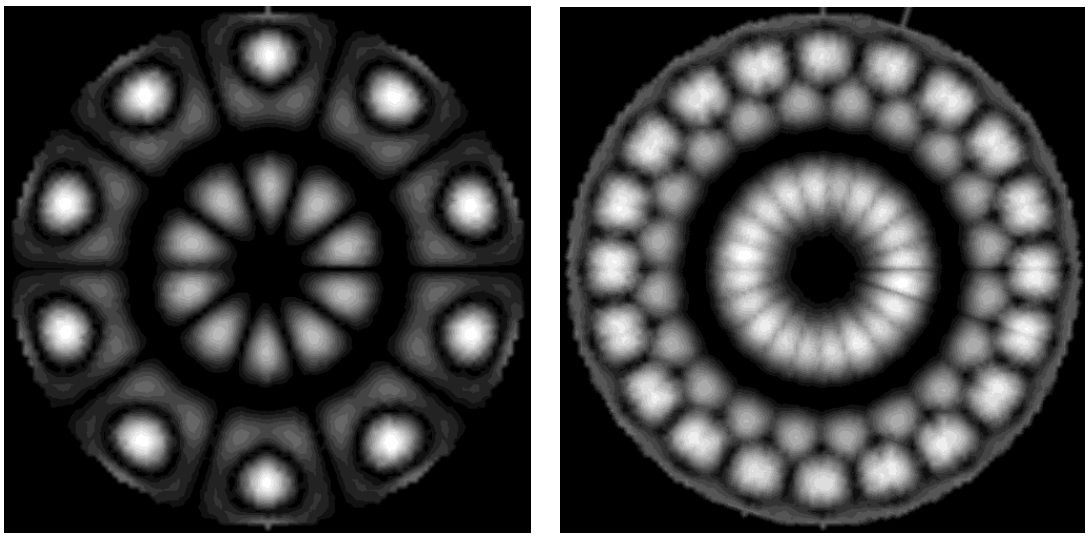


Figure 4.7: Modulation efficiency over the 500×500 mas field of view for the regular pentagon configuration (nicknamed "daisy" configuration because of the modulation map's shape). The right-hand map is deduced by superposition of two maps shifted by 18°(=360°/20). It shows an almost uniform modulation crown, about 40 mas wide.

The modulation map for the regular pentagon is shown in figure 4.7. This daisy-shaped modulation map has a big advantage but also a drawback. On one hand, the need for array rotation is reduced because a single rotation of $\pi/10$ blurs the petals of the daisy and gives a good modulation efficiency on a crown surrounding the central null. But on the other hand, even if the array is rotated, there will still remain a big crown-shaped hole of

modulation just outside the petals of the daisy. This is a serious drawback in the detection phase, since the unknown planet could be located in this big hole, and therefore not be detected. Fortunately, the use of different wavelength bins can overcome this drawback, since a variation in wavelength is equivalent to a change in the interferometer baseline. In the spectroscopic phase, the baseline should be adapted to the observation wavelength, but this is the case for all the configurations.

The pentagon configuration could therefore be a very attractive solution, if the recombination scheme was improved. For the moment, the RL2(1,0.84,0.54,0,0.70) pentagon remains our best candidate, because its modulation efficiency is two times better.

Other configurations

At least a dozen of other configurations may be found, each with a different irregular pentagonal shape. Three of these possibilities are listed below:

$$\begin{aligned}(\delta_k) &= (0, 0.781, 2.211, 3.601, 5.155)\ ,\\ (\delta_k) &= (0, 0.654, 2.618, 3.665, 5.630)\ ,\\ (\delta_k) &= (0, 1.524, 2.877, 4.385, 5.378)\ .\end{aligned}$$

All these configurations have approximately the same modulation efficiencies: mean modulation efficiency about 0.057 and maximum below 0.3. Only the modulation map significantly changes between these configurations. In fact, there remains a similitude with the daisy-shape, but some of the ten petals have a high modulation efficiency whereas other are very dark on the map (see figure 4.5 for an illustration). Therefore, these maps are not as interesting as the genuine daisy configuration because the rotations needed to get a uniform modulation crown are larger or more numerous.

Conclusions

The RL2 configurations with two FTRCs sub-arrays badly suffer from the recombination inefficiency associated with third order beam-combination. If that problem could be solved with new recombination schemes, the so-called daisy configuration could be a very interesting one, because it only uses five telescopes, and its modulation map shows interesting features.

4.2 RL configurations with six telescopes

Three RL configurations with six telescopes have already been proposed by Karlsson and Mennesson [24]. Two of them perform internal modulation between three GACs: the RL3(3,2,0,1, 0,2) configuration has already been studied in chapter 3, while RL3(1,1,0,1,1,0) will be described in section 4.2.2 as a particular case of the "1-2-4-5 family". The last one, called RL3(2,1,1,2,1,1), is composed of three sub-arrays, each made up of six telescopes. This last configuration will be briefly described in section 4.2.4.

In the following developments, we only consider RL configurations (defined in section 3.3.2), so that the diameters of all the telescopes are equal. This hypothesis has

already been justified: on one hand, it is a good way to reduce the manufacturing cost, and on the other hand, the study of 5-telescope configurations has shown that no big improvement of the modulation efficiency is obtained when using different size telescopes. We expect this property to be also valid for 6-telescope configurations to some extent. Moreover, we will discover a large number of configurations with equal size telescopes, with a variety of modulation maps, so that other configurations should not bring very different results.

4.2.1 Modulation between two GACs

By using the properties of the GACs, we show that an infinite family of 6-telescope RL configurations exists when recombining two GACs. A systematic study of such configurations is undertaken hereafter.

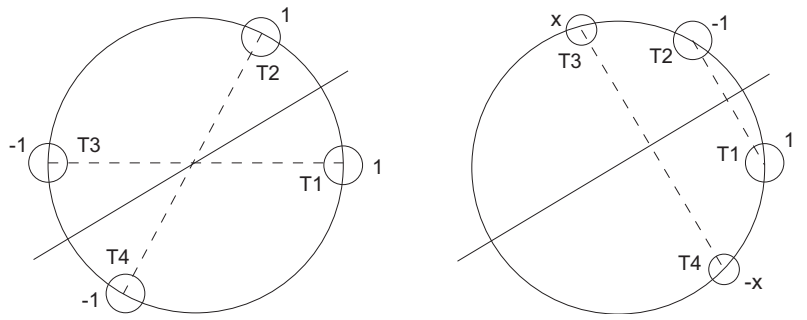
Geometrical considerations

For the superposition of two subsets of four telescopes (GACs) to form a 6-telescope interferometer, two telescopes must be common to these two sub-arrays. The other four telescopes are part of only one GAC, so that these four telescopes must all have the same diameter in their respective GAC. We will take this diameter as unity for the sake of simplicity. Let us choose the configuration of the first GAC (called GAC1), in such a way that telescopes 1 and 2 have the same size (taken as unity), and only belong to GAC1.

Taking the same notations as in appendix B.1.2 ($c_k = D_k \cos\phi_k$), the condition for an on-axis null writes $c_1 + c_2 + c_3 + c_4 = 0$. Let us see what this relation implies on c_3 and c_4. There are two possibilities to examine, depending on the relative phase shift between telescopes 1 and 2 (see figure 4.8):

- If telescopes 1 and 2 are seen with the same phase shift, arbitrarily set to zero (yielding $c_1 = c_2 = 1$), then, to ensure a completely destructive interference on the axis, the two other telescopes must satisfy $c_3 + c_4 = -2$. Now, telescopes 3 and 4 are shared with GAC2. In order to ensure that their total surface is unity as for telescopes 1 and 2, their diameter in GAC1 must be less than one. But with $|c_3| < 1$ and $|c_4| < 1$, it is impossible to satisfy the above nulling condition, and thus such a configuration does not exist.

- If telescopes 1 and 2 have a different phase shift in GAC1, then the nulling relation $c_1 + c_2 + c_3 + c_4 = 0$ becomes $c_3 + c_4 = 0$, so that telescopes 3 and 4 also have the same (and a priori arbitrary) size, and opposite phase shifts. Note that GAC1 is symmetric with respect to a diameter (figure 4.8). The size of telescopes 3 and 4 in GAC2 must be $c'_3 = -c'_4 = \sqrt{1-(c_3)^2}$ (or with an opposite sign) in order to have the same overall size for all the telescopes. The last two telescopes (number 5 and 6) are finally placed, once again symmetrically with respect to the same axis, so that, if their diameter is unity, the conditions for nulling are satisfied. There are two possibilities, one on each side of the line joining telescopes 3 and 4.



First case: $c_1 = c_2 = 1$ Second case: $c_1 = -c_2 = 1$

Figure 4.8: Generalized entrance pupil for the two cases considered above (Tk is the telescope's number in GAC1). Only the second case can give rise to an RL interferometer, when combined with one of its two "complementary" GACs.

Systematic study of this family

Let us present the successive steps needed to build an interferometer belonging to this new family:

- choose the position δ_2 of telescope 2 relative to telescope 1 (we let $\delta_1 = 0$ as usual),
- choose the position δ_3 of telescope 3. The position of telescope 4 is deduced by symmetry, and the telescope's diameters determined by the conditions to get a θ^4 on-axis extinction.
- choose a side for telescopes 5 and 6. On each side of the 3-4 line, there is a unique position for those two telescopes. Their diameter is one.

By means of an IDL code, the modulation efficiencies of a large series of configurations have been computed. The array radius, observing wavelength and field of view are the same as usual. The simulations show that a higher efficiency is reached when the telescopes are well distributed on the circle (no telescope close to another). Therefore, the best configuration is the "completely" symmetric RL2(1,1,0.707,0,0,0.707) configuration, illustrated in figure 4.9.

The modulation map is illustrated in figure 4.10, with the following values:

- Mean modulation efficiency: 0.130
- Maximum modulation efficiency: 0.654
- RMS of modulation efficiency: 0.152

Conclusion

This configuration has one of the best modulation efficiency (max and mean), but its modulation map is not uniform at all. Even with four superposed maps, we do not obtain

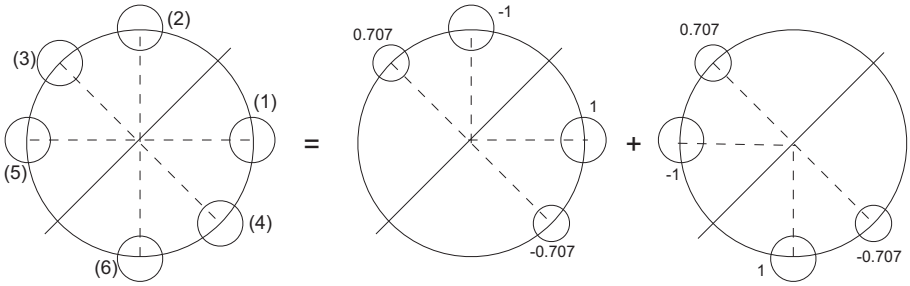


Figure 4.9: The RL2(1,1,0.707,0,0,0.707) nulling interferometer and its two sub-nulling interferometers (GACs), with their generalized entrance pupils.

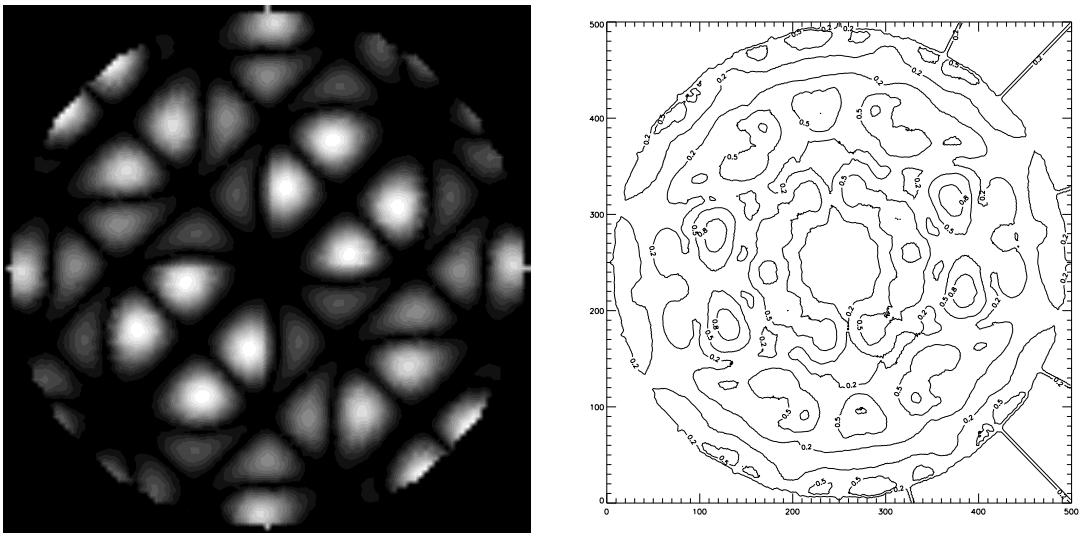


Figure 4.10: Modulation efficiency over the 500×500 mas field of view for the RL2(1,1,0.707,0,0,0.707) configuration. The right-hand map is deduced by superimposing four maps obtained by successive rotations of $\pi/8$ radians (27.5°). Isophotes at 20, 50 and 80% of maximum are shown.

a very uniform map, even if the radial directions of zero transmission are eliminated. Six rotations would be required in order to get a quite uniform modulation crown. Indeed it seems unavoidable that circular configurations with very high maximum modulation efficiency have non-uniform modulation maps (see the Karlsson pentagon for example). Nevertheless, if high modulation efficiency is desired above all, this configuration should be regarded as the best for now.

4.2.2 Modulation between three GACs

A complete study of the RL configurations with six telescopes forming three GACs cannot be easily undertaken because of the numerous degrees of freedom encountered. So, in addition to the "equal size telescopes" condition, we will restrict the study to some "regular"

configurations. What we mean by regular is that the three GACs have a similar shape. Let us distribute telescopes 1 to 6 in growing order along a circle. If GAC1 comprises telescopes 1-2-3-5, then we choose GACs 2 and 3 to be respectively composed of telescopes 3-4-5-1 and 5-6-1-3, so that a certain symmetry is preserved in the GACs' arrangement. This symmetry argument can be justified because the previous developments tend to prove that the configurations preserving some symmetry either have the best modulation efficiencies, or show interesting features in their modulation map.

Restricting to those regular configurations, we have three cases to consider, depending on the choice for the first GAC. The three possibilities are 1-2-3-4, 1-2-3-5 and 1-2-4-5. All other configurations can be reduced to one of these by re-labeling the telescopes.

Note that except for the regularity of the GACs' pattern, no real symmetry condition is imposed on the geometry of the array. But we will see below that the regularity of the pattern implies itself some symmetry in the geometrical configuration of the six telescopes (which is indeed quite a good news).

The 1-2-3-4 configurations

The geometry for a 1-2-3-4 configuration is illustrated in figure 4.11. GACs 2 and 3 are respectively composed of telescopes 3-4-5-6 and 5-6-1-2. By using the nulling equations for the three GACs, we prove below that the only allowed geometry is a superposition of two equilateral triangles, with an arbitrary angle between them. An angle of 40° is chosen in figure 4.11.

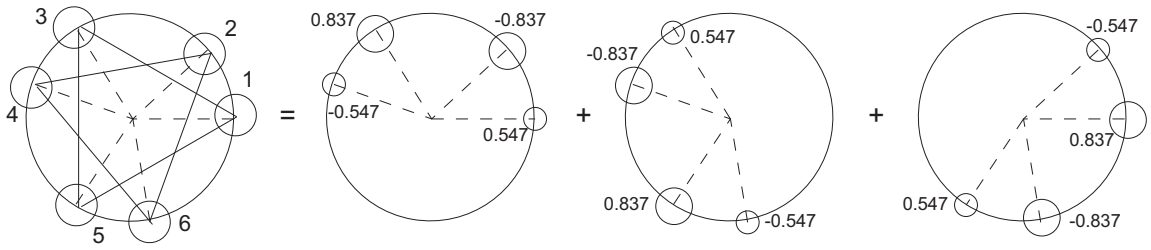


Figure 4.11: An hexagonal 1-2-3-4 nulling interferometer and its three sub-nulling interferometers (GACs), with their generalized entrance pupils. The six telescopes are forming two equilateral triangles, with a relative angle of 40° in this example.

In order to build a 1-2-3-4 configuration, one has to take the following steps:

- The first GAC can be arbitrarily chosen. Both the diameters and the positions of the telescopes are a priori free. So, a geometry $(\delta_1, \delta_2, \delta_3, \delta_4)$ is chosen and the generalized diameters c_1, c_2, c_3, c_4 are deduced to form a GAC.

- GAC2 shares telescopes 3 and 4 with the first GAC. These telescopes are numbered 1' and 2' in GAC2. In order to equalize the telescopes' size, we impose a relation between c'_1 and c'_2: $(c_3)^2 + (c'_1)^2 = (c_4)^2 + (c'_2)^2$. The positions δ'_1 and δ'_2 are already fixed (equal to δ_3 and δ_4), so that five degrees of freedom remain for GAC2 (c'_1, c'_3,

c'_4 and δ'_3, δ'_4) whereas only three nulling equations are to be satisfied. Therefore, we can arbitrarily choose the two remaining positions (δ'_3 and δ'_4), and the generalized diameters are deduced from the nulling equations.

- The last GAC shares its telescopes 1" and 2" with telescopes 3' and 4' of GAC2, and its telescopes 3" and 4" with telescopes 1 and 2 of GAC1. So, the diameters in GAC3 can be deduced from the "equal size telescopes" conditions. Moreover, all the positions are known. Only the phase shifts (i.e., the sign of the generalized diameters c''_k) are unknown. The only thing to do with GAC3 is therefore to see whether these parameters are actually forming a GAC, i.e., whether they satisfy the nulling equation for a θ^{-4} rejection rate.

The previous steps have been implemented in an IDL code, with the following result: the nulling equations for GAC3 are satisfied if and only if the six telescopes are arranged on two equilateral triangles inscribed on the circle. In that particular case, the three GACs are actually the same (to within a 120° rotation, see figure 4.11). Thus, a single parameter (the angle between the two triangles) is sufficient to define a 1-2-3-4 RL3 configuration.

Modulation efficiencies have been computed for a large series of configurations (by varying the angular parameter). The best modulation characteristics are obtained with a totally symmetric configuration: a regular hexagon. It is the RL3(2,2,1,0,0,1) configuration. Its modulation characteristics are:

- Mean modulation efficiency: 0.128
- Maximum modulation efficiency: 0.340
- RMS of modulation efficiency: 0.108

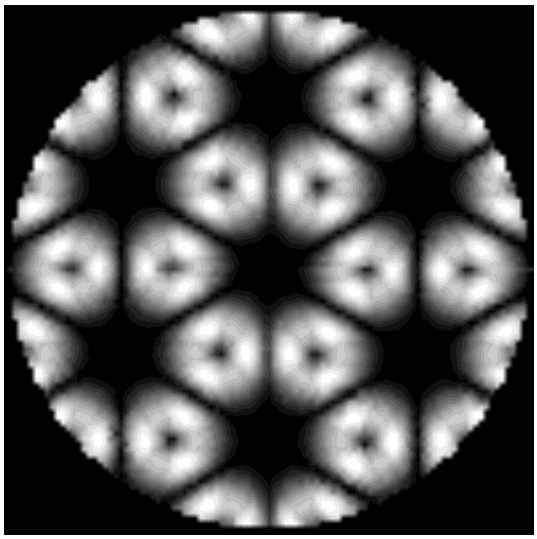


Figure 4.12: Modulation map for the RL3(2,2,1,0,0,1) configuration.

The modulation map, illustrated in figure 4.12 is exactly the same as the map for the hexagonal RL3(3,2,0,1,0,2) configuration already studied in section 3.3.2. This property has already been encountered in the 5-telescope configurations: any given array geometry can give the same modulation characteristics with different recombining schemes. Note that in this case, the efficiencies are slightly better than in the RL3(3,2,0,1,0,2) case.

The 1-2-3-5 configurations

The procedure to build a 1-2-3-5 configuration (illustrated in figure 4.13) is straightforward: since telescopes 2, 4 and 6 only belong to a single GAC, they are all taken with a unitary diameter, in order to satisfy the "equal size telescopes" condition. Then, if the geometry is chosen, there remain 3×3=9 unknown parameters (the three remaining telescope diameters in each GAC) for 9 nulling equations. So, for a given geometry, there is only one nulling configuration. But this configuration will be a priori composed of different size telescopes. The last thing to do is therefore to verify the "equal size telescopes" conditions for telescopes 1, 3 and 5. With the same notations as above, these conditions read:

$$\begin{aligned} (c_1)^2 + (c'_4)^2 + (c''_3)^2 &= 1\,, \\ (c_3)^2 + (c'_1)^2 + (c''_4)^2 &= 1\,, \\ (c_4)^2 + (c'_3)^2 + (c''_1)^2 &= 1\,. \end{aligned}$$

Using an IDL code and exploring a large series of geometrical configurations, we conclude that, just as for the 1-2-3-4 configurations, the only allowed geometry is the superposition of two equilateral triangles (see figure 4.13). All other geometrical configurations yield different size telescopes.

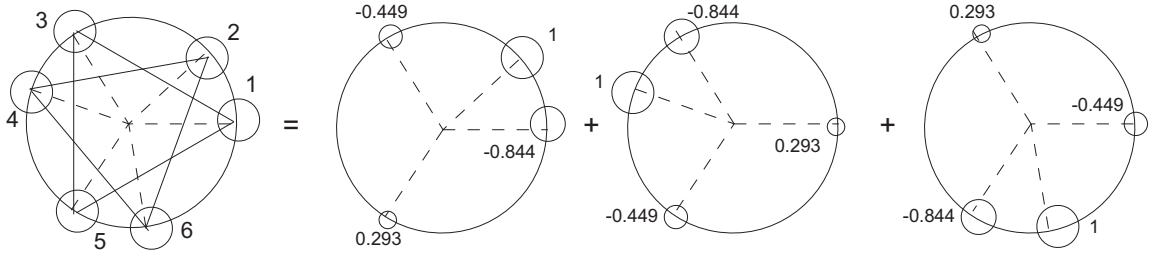


Figure 4.13: An hexagonal 1-2-3-5 nulling interferometer and its three sub-nulling interferometers (GACs), with their generalized entrance pupils. The six telescopes are forming two equilateral triangles, with a relative angle of 40° in this example.

As in the previous case, we only have to vary a single parameter (the angle between the two triangles) to explore the modulation characteristics of all configurations. As expected, the regular hexagon (i.e., the most symmetrical configuration) yields the higher *mean* modulation efficiency:

- Mean modulation efficiency: 0.121
- Maximum modulation efficiency: 0.299

- RMS of modulation efficiency: 0.102

This configuration is indeed the RL3(3,2,0,1,0,2) regular hexagon, first proposed by Karlsson and presented in section 3.3.2.

But unlike the previous case, this configuration has not the highest *maximum* modulation efficiency: when the angle between the two triangles is equal to 40°, the modulation characteristics are:

- Mean modulation efficiency: 0.110
- Maximum modulation efficiency: 0.416
- RMS of modulation efficiency: 0.084

This is the RL3(1,0.449,0,0.293,0,0.844) configuration, illustrated in figure 4.13. Its modulation map is shown in figure 4.14. Surprisingly, it has a perfect six-fold symmetry as the regular hexagon. As a consequence, the superposition of three maps shifted by 20° with respect to each other produces a very uniform sky coverage: the modulation efficiency is larger than 20% almost everywhere over a 140 mas-wide crown! It is therefore very well suited to the detection phase, even better than the regular hexagon.

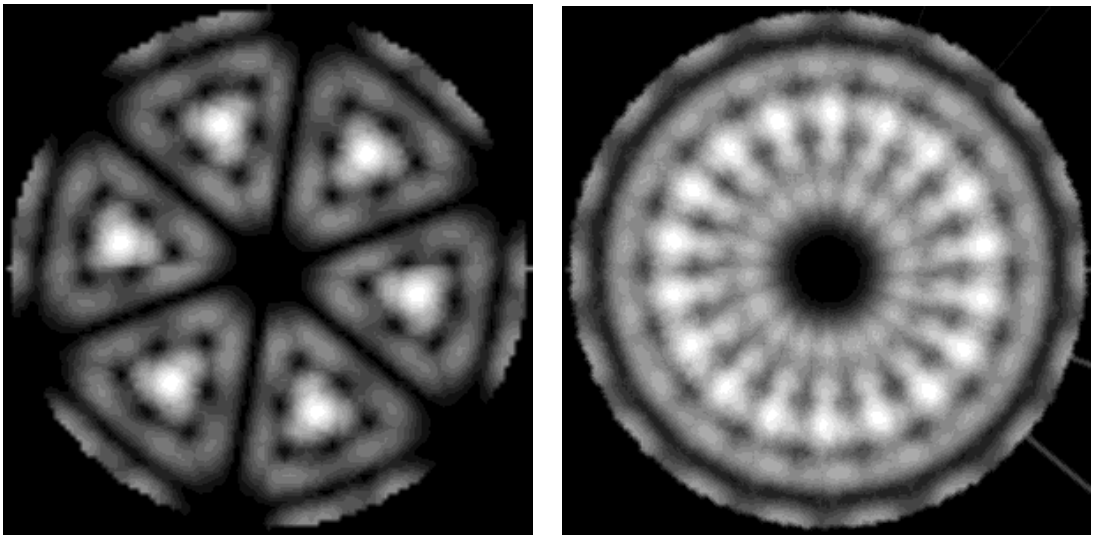


Figure 4.14: Modulation efficiency over the 500×500 mas field of view for the RL3(1,0.449,0,0.293,0,0.844) configuration. The right-hand map is deduced by superimposing three individual maps obtained by successive rotations of 20°.

The 1-2-4-5 configurations

In order to build a 1-2-4-5 configuration, the same procedure as for a 1-2-3-4 configuration can be used, because any two GACs have two telescopes in common. The only difference comes from the telescope labeling. Once again, the numerical simulations show that the

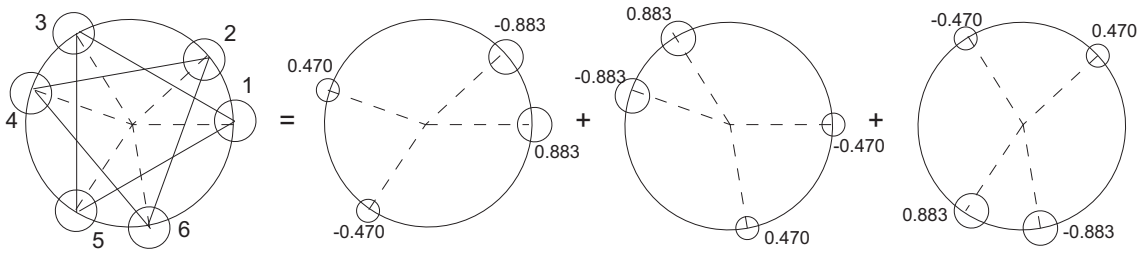


Figure 4.15: An hexagonal 1-2-4-5 nulling interferometer and its three sub-nulling interferometers (GACs), with their generalized entrance pupils. The six telescopes are forming two equilateral triangles, with a relative angle of 40° in this example.

only allowed geometry consists in the superposition of two equilateral triangles (figure 4.15).

The regular hexagon RL3(1,1,0,1,1,0) seems apparently very attractive, since all telescopes participating to a GAC are seen with the same amplitude, and therefore only 50:50 beam-splitters are used. But unfortunately, its modulation characteristics are very bad. Indeed, writing down the expressions for the amplitude of each GAC, we immediately see that these amplitudes are real, so that the modulation efficiency is null according to equation (3.14). This fact has already been stated by Mennesson [38].

All other possibilities for the angular parameter have been investigated. The best modulation efficiency happens for a 40° angle between the two triangles:

- Mean modulation efficiency: 0.088
- Maximum modulation efficiency: 0.589
- RMS of modulation efficiency: 0.099

This is the RL3(1,1,0,0.532,0.532,0) configuration, which has one of the highest maximum modulation efficiency, but quite a low mean efficiency. It would therefore be well suited for spectroscopic investigations, when the planetary location is known, but not very efficient in the detection phase. Its modulation map is shown in figure 4.16. The six modulation maxima are quite narrow. Therefore, a large number of rotations (at least six) are needed to get a uniform modulation map.

Conclusion

In this section, a large number of new configurations were found, and the two known configurations re-discovered. These facts can be taken as an a posteriori justification of the "regularity" hypothesis we made at the beginning: if only a small number of configurations had been found, we would have been obliged to go back and make other hypotheses. The modulation efficiencies of these new configurations are not significantly higher than the one we found for the Karlsson pentagon. The mean modulation efficiency can almost reach 0.13, but the maximum efficiency is generally not very high (except for the last configuration).

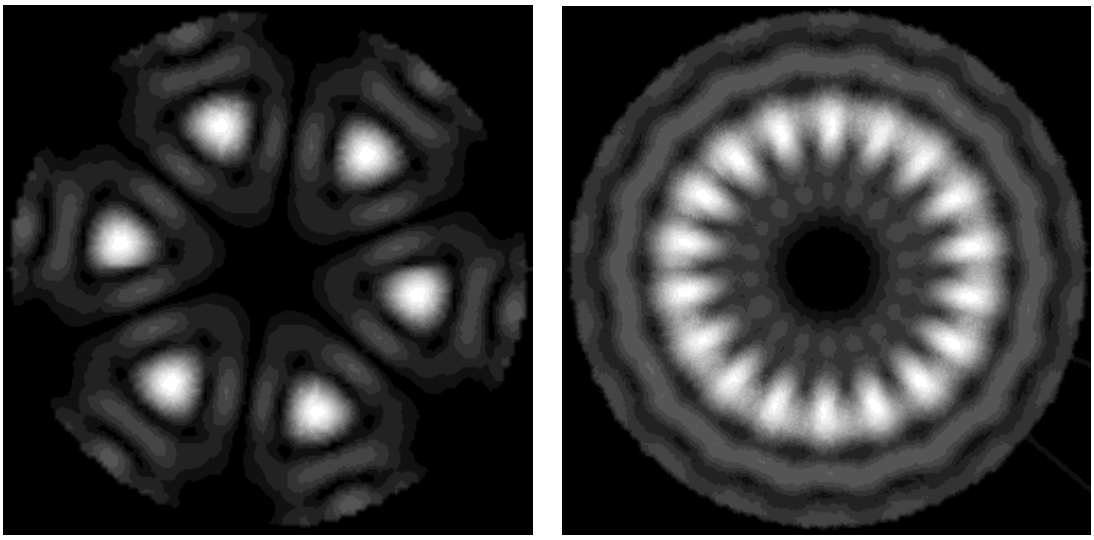


Figure 4.16: Modulation efficiency over the 500×500 mas field of view for the RL3(1,1,0,0.532,0.532,0) configuration. The right-hand map is deduced by superimposing three maps obtained by successive rotations of 20°. Note that the modulation crown is not uniform at all. Some radial directions have a modulation efficiency smaller than 10%, while the bright maxima have a modulation efficiency larger than 50%. More than three rotations would therefore be required in the detection phase.

This fact is easily understandable if we go back to the modulation efficiency definition (3.21) for an RL3 configuration:

$$\frac{|S_{12} - S_{21}| + |S_{13} - S_{31}| + |S_{23} - S_{32}|}{6I_0} \,. \tag{4.4}$$

Each $|S_{ij} - S_{ji}|$ term in this sum represents the useful planetary part of the signal for a *i*-*j* pair of GACs. It is quite obvious that the maxima of modulation efficiency between GACs 1 and 2 will not happen at the same places on the sky as for the 1-3 and the 2-3 pairs. So, even if the modulation efficiency is locally very high for a pair of GACs, the global efficiency for the three GACs will not generally be so high because only a third of the total signal ($6I_0$) is directed to this output, the two other thirds being directed towards the two other pairs of GACs (with a lowest efficiency). This is the reason why the 3-GAC configurations are less efficient when the planet is located (spectroscopy phase).

However, a special attention must be given to the 40° geometry. When used in the 1-2-3-5 recombination scheme, this arrangement gives both a good mean and a good maximum efficiency. Moreover, as proposed by Lawson *et al.* in [29], it could be possible to design TPF (or Darwin in our case) with the ability to reconfigure its beam-combining optics through an exchange of combiner modules. This ability could be used in order to optimize the instrument for each phase: keeping the same array geometry, the 1-2-3-5 configuration (with a good mean efficiency) would be used during the detection phase, and the 1-2-4-5 configuration during the spectroscopy phase, when a high maximum efficiency is required[7].

[7]During the spectroscopic study of a planet, since the planetary location is known, the baseline is

4.2.3 Modulation between FTRCs

The study of those configurations is a very tedious work, since there are many degrees of freedom to take into account. In fact, this study is not very useful for two reasons:

- The main reason comes from the recombining scheme for the 5-telescope sub-arrays (FTRCs). We have proven in section 4.1.2 that a third order recombination scheme is necessary when sub-arrays composed of five telescopes are involved (whereas a second order scheme is sufficient for GACs). The addition of a third recombining level reduces the recombination efficiency by a factor of two, so that the modulation efficiency is about two times lower than that of GACs.

- Even if a better recombination scheme could be found, there should not be a great advantage in using FTRCs instead of GACs. On one hand, this complicates the recombination optics by involving more telescopes in each sub-array, and on the other hand, a great variety of modulation maps are already available when GACs are recombined.

For these reasons, we will not undertake a systematic study of RL configurations with six telescopes performing internal modulation between FTRC sub-arrays.

4.2.4 Modulation between STRCs

The abbreviation "STRC" stands for a "Six Telescopes Real Circular" configuration, yielding a θ^4 on-axis extinction (same notation as for FTRC). For the same reason as above (i.e., the need for third order recombination schemes), internal modulation between STRCs should not be very efficient. We will anyway say a word about one of these configurations, the RL3(2,1,1,2,1,1), which has been proposed by A. Karlsson in [24]. Its regular hexagonal geometry is illustrated in figure 4.17.

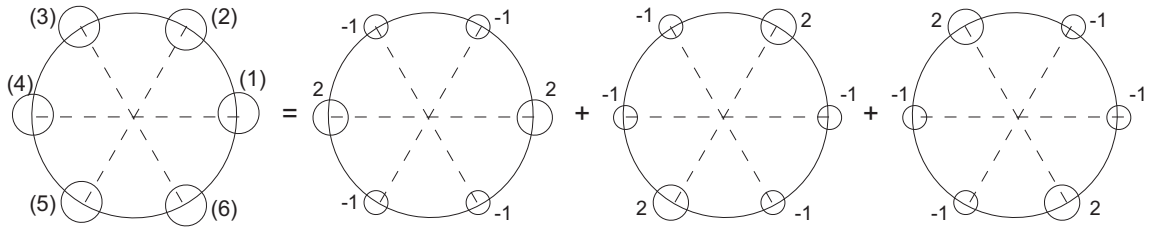


Figure 4.17: The RL3(2,1,1,2,1,1) nulling interferometer and its three sub-nulling interferometers (STRCs), with their generalized entrance pupils.

As expected, its modulation efficiency is far weaker than those for the other configurations:

- Mean modulation efficiency: 0.029

optimized for that location, with a maximum modulation permanently pointed towards that direction.

- Maximum modulation efficiency: 0.113
- RMS of modulation efficiency: 0.024

This configuration is therefore of no practical interest.

Another remark about internal modulation between STRCs is that a θ^{-6} rejection rate is not reachable. Indeed, it can easily be proven that, for any given circular geometry, there exists one and only one real nulling configuration yielding a θ^{-6} rejection rate. Since two sub-arrays are needed to perform internal modulation, this rejection rate is not achievable when internal modulation is used.

4.3 Linear configurations with internal modulation

4.3.1 General considerations

We have seen in section 3.3.3 how two linear sub-arrays (Bracewell, DAC or even OASES) can be recombined to form a linear array with internal modulation. Some general properties of these configurations can be deduced without any calculation, and are discussed below.

The most evident property of linear configurations is that the maxima of the modulation maps always have a linear shape, the lines of maximum efficiency being perpendicular to the array. This property comes from the fact that the two sub-arrays are on a same line, so that their linear fringe patterns are parallel to each other. This is in fact one serious drawback for linear configurations, since a uniform modulation map is required. In order to get a mean modulation map with sufficient uniformity, the array needs to be rotated at least three or four times, with large amplitudes, which consumes propellant and time.

A second feature of linear arrays is the need to equalize the optical paths between the telescopes and the recombination hub, in order to avoid long delay lines. Since no point is equidistant from the telescopes, reflections must be introduced on the optical paths to lengthen the shortest of them. If the distances between successive telescopes are equal, this can be done by sending the beams to neighboring telescopes before sending them to the beam-combiner (see figure 4.18). However, by routing the starlight from one telescope to another, the control laws of the telescopes may be overly constrained (see [29]). When the distances between successive telescopes are not equal any more, delay compensation becomes a complicated task, and simple solutions do not often exist. Optical delay compensation is therefore a serious limitation to linear arrays.

With these two drawbacks, why then use linear arrays? The answer comes from the number of telescopes: internal modulation can be achieved with only four telescopes, still ensuring a θ^{-4} rejection rate. Moreover, the sub-nulling interferometers being composed of three or four telescopes, second order recombination schemes are used (see section 4.1.2), giving a good recombination efficiency with quite simple optics.

In this section, the most interesting configurations (i.e., with the smallest possible number of telescopes) of section 3.3.3 are studied, as well as other types of configurations, involving non regular DACs, or recombination of DACs in a two dimensional scheme.

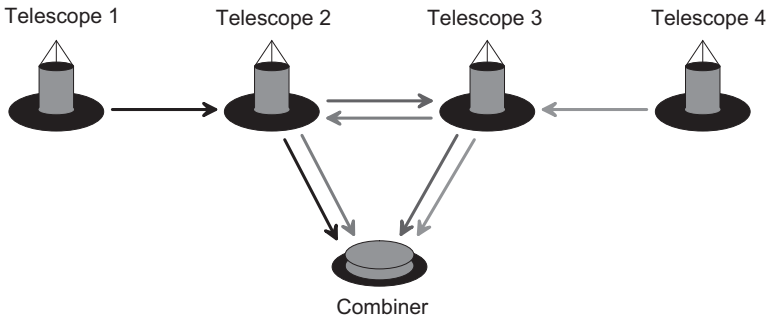


Figure 4.18: Delay compensation in the plane of the array (taken from [29]). This method is currently envisaged for TPF.

4.3.2 Modulation between two regular DACs

Linear configurations

There are three possible ways to recombine two DACs in a linear configuration, depending on the number of telescopes they share (see section 3.3.3). Among those possibilities, the most interesting one is clearly the 4-telescope linear array because the number of telescopes is minimum. Let us study this configuration in detail.

The effective diameters for a regular DAC being in a $1{:}\sqrt{2}{:}1$ ratio, the diameters for the global array are in a $1{:}\sqrt{3}{:}\sqrt{3}{:}1$ ratio (see figure 4.19). For the study, a distance of 10 m between the telescopes is chosen, so that the total baseline is 30 m, allowing an easy comparison with the circular configurations (30 m in diameter).

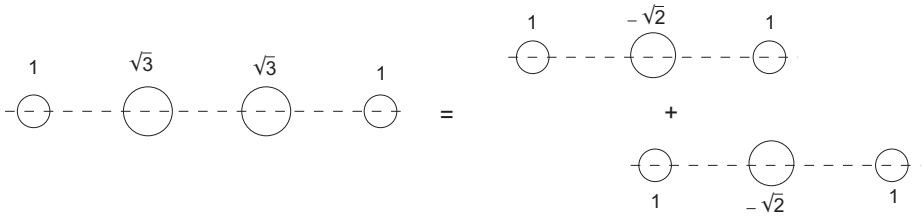


Figure 4.19: The 4-telescope linear array with internal modulation between two regular DACs. Note the $1{:}\sqrt{3}{:}\sqrt{3}{:}1$ ratio for the total diameters.

The modulation characteristics of this configuration are listed below:

- Mean modulation efficiency: 0.181
- Maximum modulation efficiency: 0.518
- RMS of modulation efficiency: 0.191

Compared to circular configurations, this is a very good modulation efficiency. The modulation map is shown in figure 4.20. The parallel lines of maximum modulation are a characteristic of linear configurations. In order to make the modulation map more uniform, two successive rotations of 60° are applied to the array, and the three maps superimposed (figure 4.21). The resulting map shows a bright modulation crown surrounding the central null, with a good uniformity: the relative efficiency is above 50% of maximum almost everywhere (i.e., modulation efficiency above 25%). An almost complete sky coverage can even be achieved with four superposed maps (figure 4.22).

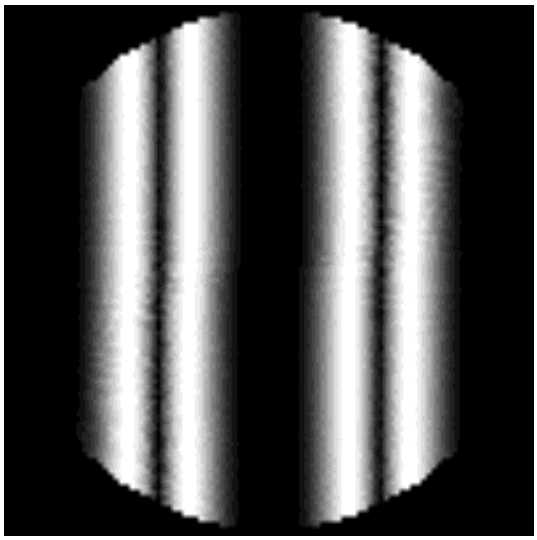


Figure 4.20: Modulation efficiency over the 500×500 mas field of view for the 4-telescope regular linear array.

Let us sum up the advantages and drawbacks of this first linear configuration:

- High modulation efficiency (mean and max)
- Only four telescopes, but of different sizes
- Non uniform modulation map (composed of parallel lines)
- Need for large and numerous rotations to make the modulation map uniform
- Need for beam reflection on a neighboring telescope to equalize the optical paths

A quantitative study of the practical consequences of these advantages and drawbacks should be one of the next goals for the Darwin team.

A last remark about this configuration concerns delay compensation for the beams. We have seen in figure 4.18 that a simple way to compensate for the optical path difference between the beams is to use reflections on neighboring telescopes. Another proposal is described in [29]. It uses a non plane configuration with a parabolic shape. This would

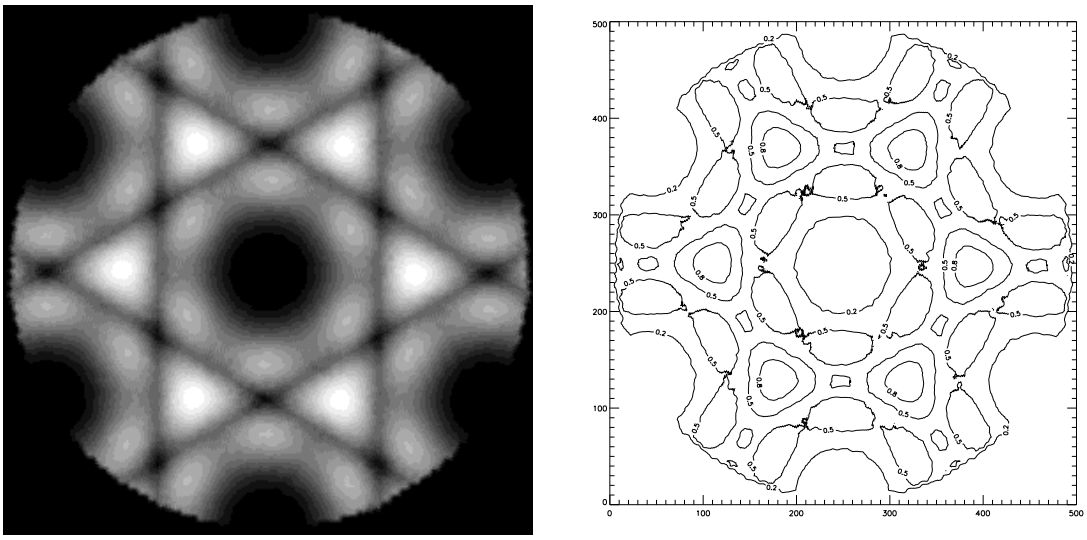


Figure 4.21: This map is deduced from figure 4.20 by superimposing three maps obtained by successive rotations of 60°. Isophotes at 20, 50 and 80% of maximum are shown.

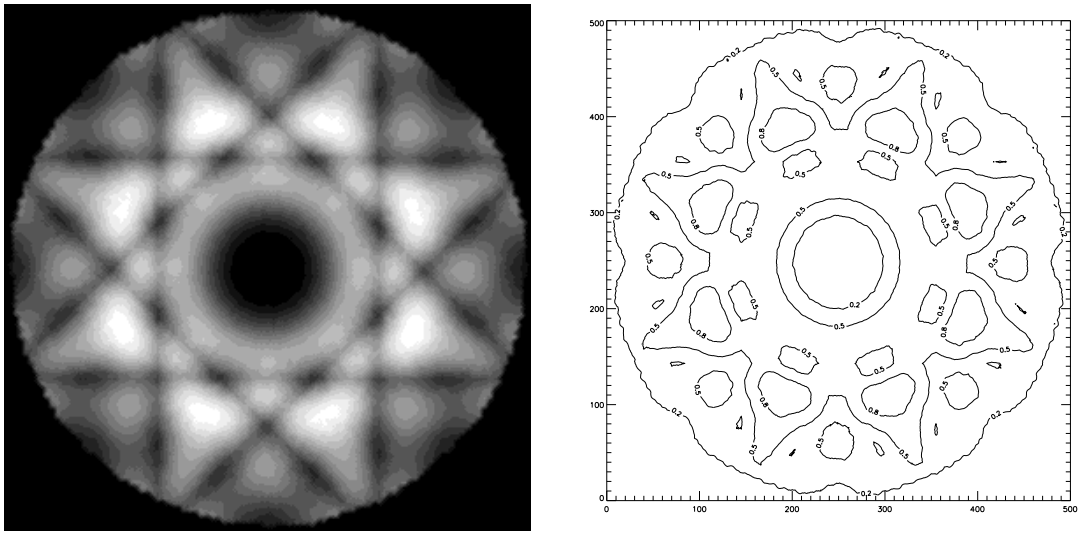


Figure 4.22: This map is deduced from figure 4.20 by superimposing four maps obtained by successive rotations of 45°. Isophotes at 20, 50 and 80% of maximum are shown.

have been a good solution if the thermal coupling between the telescopes was not a serious problem. We have shown in section 3.3.1 that non plane configurations cannot be accepted.

After dealing with the 4-telescope linear array, the recombination of two DACs sharing only one telescope should also be considered, since five telescopes are still a low-cost solution, even if the telescopes do not have the same size. In this case, the diameters are in a $1{:}\sqrt{2}{:}\sqrt{2}{:}\sqrt{2}{:}1$ ratio. In order to keep approximately the same angular resolution as

before, the length of each DAC remains 20 m, so that the total length of the array is 40 m.

- Mean modulation efficiency: 0.201
- Maximum modulation efficiency: 0.778
- RMS of modulation efficiency: 0.268

This is indeed quite better than any other configuration. A maximum efficiency of 78% is probably one of the best reachable efficiencies. So, even if five telescopes of different sizes are used, and even if large rotations are needed to make the map uniform, this configuration has an obvious interest.

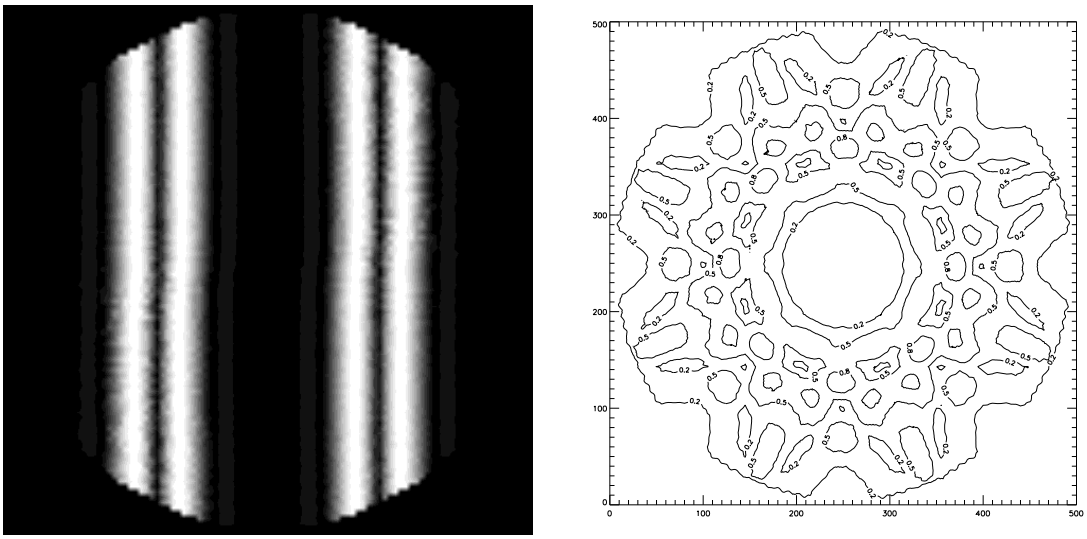


Figure 4.23: Modulation efficiency over the 500×500 mas field of view for the 5-telescope regular linear array. The right-hand map is deduced by superimposing four maps obtained by successive rotations of 45°. Isophotes at 20, 50 and 80% of maximum are shown.

The modulation map is shown in figure 4.23, together with the superposition of four rotated maps. Just as in the previous case, the superposition of three maps is sufficient to get an almost uniform modulation crown, but an even better sky coverage is obtained with four maps: a relative efficiency above 50% of maximum is maintained over a crown about 30 mas wide. This means that a planet located between 80 and 110 mas from the central star is modulated with an efficiency better than 40% when the observing wavelength is 10 μm.

Unfortunately, there is no obvious way to compensate the OPDs between the beams for this 5-telescope linear configuration. If a simple way to achieve the equality of the optical paths could be found, this configuration would surely be one of the most interesting because of its very high maximum modulation efficiency.

Two dimensional configurations

When the two regular DACs are not on a same line any more, they cannot share more than one telescope. There are three different ways for two DACs to share a telescope (see figure 4.24). Let us briefly describe those three configurations. In order to allow an easy comparison between them, we will adjust the scale of each array so that the maximum distance between two telescopes is 30 m. The field of view is still 500 mas in diameter, and the observing wavelength 10 μm.

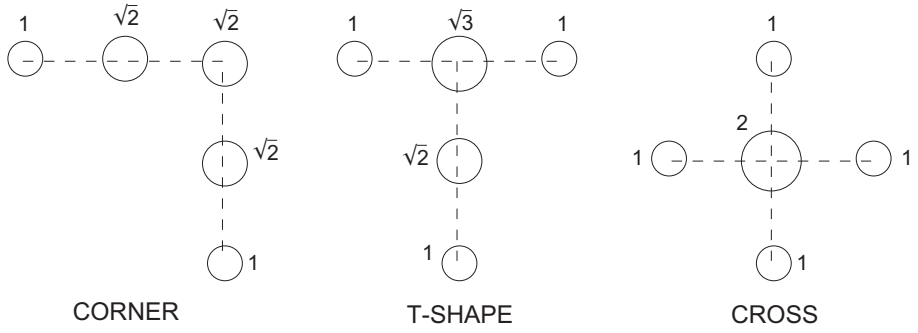


Figure 4.24: The three possible ways to recombine two regular DACs in a two dimensional array. Note that the angle between the two DACs can be different from 90°.

- **Corner:** The distance between two successive telescopes in a DAC is set to 10.6 m. This ensures that the distance between the two ends is about 30 m (with an angle of 90° between the two DACs). The computed modulation efficiency has a maximum of 0.547 and a mean of 0.081. The mean efficiency is low because the maxima are quite narrow: the modulation map is dominated by dark regions (with low modulation efficiency) and is not uniform at all (see figure 4.25). This configuration is therefore not interesting. These properties still hold when the angle between the two DACs is modified.

- **T-shape:** The distance between two successive telescopes in a DAC is set to 13.4 m, and the angle to 90° as illustrated. The computed modulation efficiency has a maximum of 0.448 and a mean of 0.083. Once again, the modulation map is not very uniform (see figure 4.25), and the uniformity cannot be easily improved with a small number of rotations. Moreover, the modulation efficiency is low when compared with linear arrays.

- **Cross:** For this configuration, some analytical calculations easily show that the output of both DACs are real, so that the modulation efficiency is zero. The numerical simulations confirm this analytical result.

Obviously, two dimensional configurations with internal modulation between two DACs are far less interesting than linear configurations.

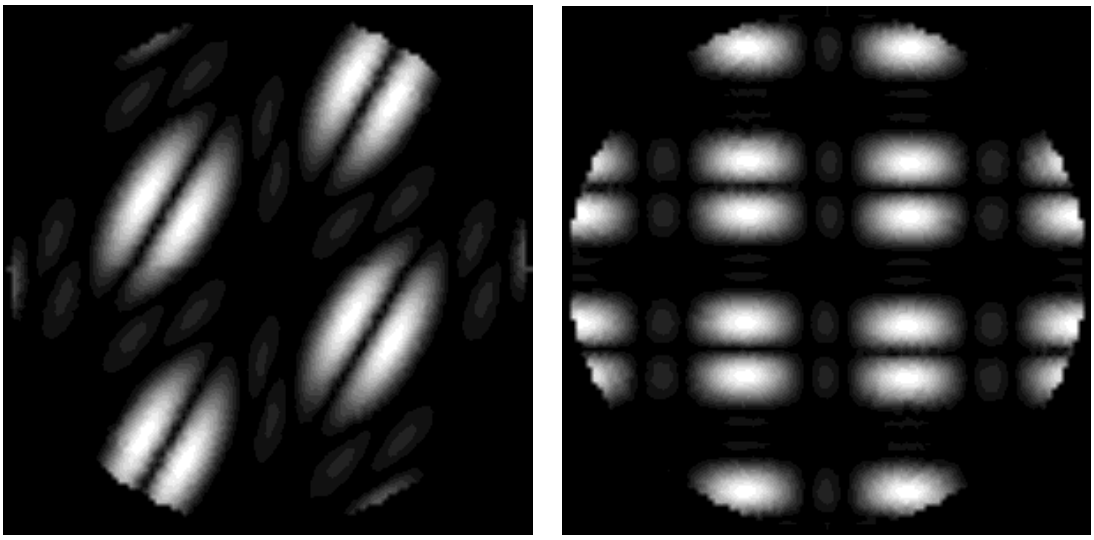


Figure 4.25: Modulation efficiency over the 500×500 mas field of view for the corner configuration (left) and the T-shaped configuration (right). Both maps are not uniform at all.

4.3.3 Modulation between two irregular DACs

Instead of recombining two regular DACs, we can use any member of the DAC family (presented in section 3.2.2). We will restrict to linear configurations composed of four telescopes, because, on one hand, linear configurations with five telescopes do not allow simple delay compensation, and, on the other hand, two dimensional configurations do not have very good modulation characteristics.

By using non regular DACs, we should try to eliminate one of the drawbacks of the 4-telescope regular array. For instance, we could try to get equal size telescopes, or to reduce the number of reflections needed to equalize the OPDs. On the other hand, the uniformity of the map cannot be improved, since all linear configurations have the same kind of modulation map, with parallel bright and dark lines.

Equal size telescopes

Even if the central telescope of a single DAC is theoretically the biggest, its effective diameter is divided by $\sqrt{2}$ (remember figure 3.8). Therefore, the central telescope is not necessarily the biggest. This property can be used to keep four telescopes with equal diameters when superimposing two irregular DACs. A 3:$\frac{4}{\sqrt{2}}$:1 diameter ratio can achieve this equality since $9 = 16/2 + 1$ (see figure 4.26).

As in the regular case, the modulation efficiency of this linear configuration is very good. In order to have a 30 m array, we have chosen $L = 6$ m (see figure). With this value, the modulation characteristics are:

- Mean modulation efficiency: 0.184

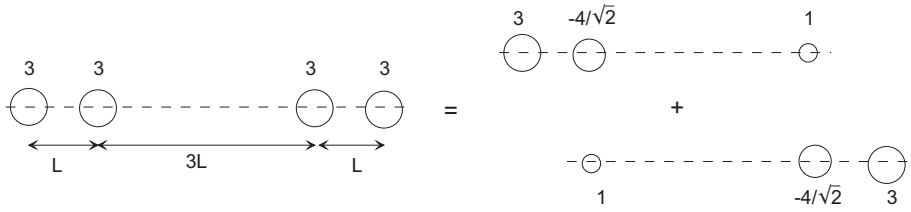


Figure 4.26: The equal-size-telescope linear array, with its two sub-nulling interferometers (irregular DACs). The distances between successive telescopes are determined in each DAC by the relation $D_1L_1 = D_3L_3$, if L_1 and L_3 are the distances between the central telescope and the other two in this particular DAC.

- Maximum modulation efficiency: 0.673
- RMS of modulation efficiency: 0.190

Unfortunately, the modulation map shows that the first two maxima are neither large nor bright (see figure 4.27). This has a bad influence on the detection phase since the modulation crown surrounding the central null is proportional to the first maximum. Superposing three maps, we obtain a modulation crown about 25 mas wide with a modulation efficiency hardly larger than 20% of maximum (i.e., efficiency superior to 14%). This is not a very good result, since even the Karlsson pentagon has better characteristics. The only possibility to improve the modulation efficiency in the detection phase is to use one of the brightest maxima, but a large number of rotations is necessary in order to make it uniform. Note that, even if this configuration is less efficient than the 4-telescope regular array in the detection phase, it has a better efficiency in the spectroscopic phase, because a maximum efficiency of 67% can be reached.

The last thing to deal with is delay compensation. Since reflections on neighboring telescopes do not seem to allow simple delay compensation, a new possibility is considered. We have not spoken about the metrology hub yet. This satellite, unlike the recombination hub, has to be "out-of-plane", on the Sun-side of the telescopes, in order to control the telescopes positions with a high accuracy. Mirrors could be mounted on it in order to equalize the optical paths, as shown in figure 4.28. Note that the telescope satellites will heat up the metrology hub (due to reflected light and to thermal emission of the sunshields), but the mirrors mounted on it could probably be kept cool.

Other possibilities

When recombining two irregular DACs in a 4-telescope linear array, there are two types of configurations to consider, depending on the relative distances between neighboring telescopes. The first type corresponds to the case where the distance between telescopes 2 and 3 (inner telescopes) is larger than the distance L between telescopes 1 and 2 (or

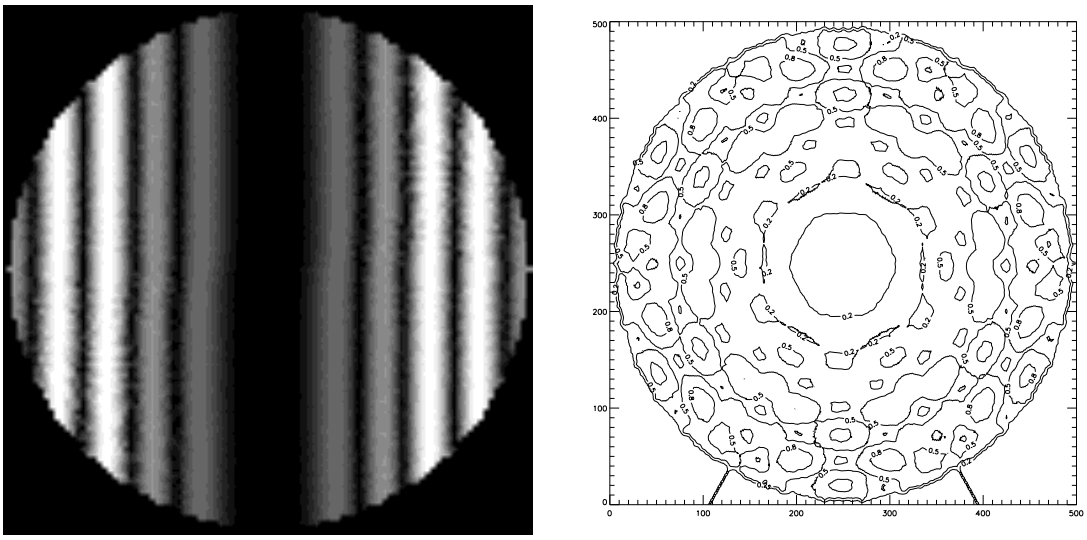


Figure 4.27: Modulation efficiency over the 500×500 mas field of view for the 1:1:1:1 linear configuration. The right-hand map is deduced by superimposing three maps obtained by successive rotations of 60°. Isophotes at 20, 50 and 80% of maximum are shown.

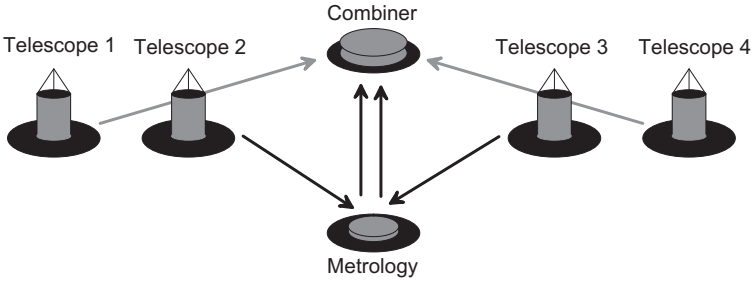


Figure 4.28: Delay compensation by using reflections on the out-of-plane metrology hub.

telescopes 3 and 4), and the second type to the other case. The modulation maps of these two types of configurations have different properties. We will compare them to the modulation map of the regular array, which has two pairs of bright lines, each having the same modulation efficiency (see figure 4.20).

- *Type 1.* When the distance between inner telescopes is larger than L, the two contiguous bright lines in figure 4.20 do not have the same modulation efficiency any more. The inner lines become fainter than the outer ones. This has already been proven in the previous case with figure 4.27. This property has a bad influence on the detection phase as proven above. But, on the other hand, the maximum modulation efficiency grows significantly, which is an advantage in the spectroscopic phase. The same delay compensation (figure 4.28) can be applied to all configurations of this type.

- *Type 2.* When the distance between inner telescopes is smaller than L, the inner lines in figure 4.20 become brighter and larger than the outer lines. This improves the sky coverage in the detection phase since the first modulation crown becomes large and bright when three maps are superposed. But unfortunately, the maximum efficiency rapidly decreases with decreasing distance between inner telescopes. For example, when the distance ratio is 1:2, the maximum modulation efficiency is only 0.326. So, the improvement in the map's uniformity is balanced by the decrease in efficiency, and there is no real improvement indeed. Moreover, the decrease in efficiency has a very bad influence for the spectroscopic phase: instead of a 51.8% efficiency, the planetary signal is only modulated with a poor 32.6% efficiency. Note that delay compensation could be somewhat simplified with a type 2 configuration, since the beams coming from the outer telescopes do not have to be reflected on an inner telescope any more (see figure 4.29).

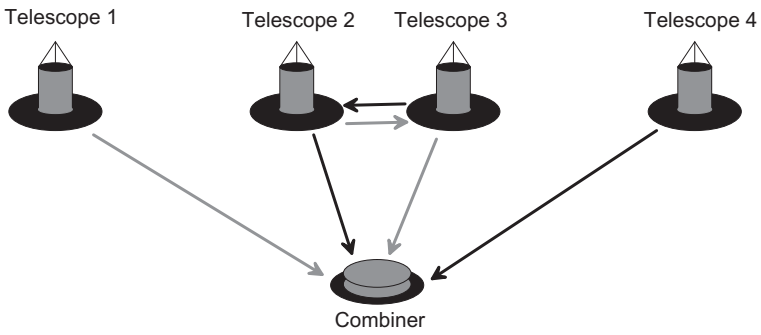


Figure 4.29: Delay compensation in the plane of the array for a "type 2" linear configuration, for which the distance between inner telescopes is smaller than L.

From these properties, we conclude that the regular linear array (with equidistant telescopes) is probably the best compromise between a good sky coverage in the detection phase and a high modulation efficiency in the spectroscopic phase.

4.3.4 Modulation between OASES configurations

Internal modulation between two OASES configurations[8] could also be interesting, since only five telescopes are necessary if three telescopes are shared between the two OASES. But there would be three different sizes for the telescopes (global diameters in a $1{:}\sqrt{2}{:}2{:}\sqrt{2}{:}1$ ratio), and the numerical simulations show that the modulation efficiency is not better than when DACs are recombined. These are the reasons why internal modulation between two OASES configurations will not be considered. Note that delay compensation would also constitute a major difficulty.

[8] An OASES configuration is a linear configuration with four telescopes, giving a θ^6 starlight suppression (see appendix C.2.2).

4.3.5 Conclusion

Obviously, the characteristics of linear configurations with internal modulation are quite different from those of circular configurations. The modulation efficiency and sky coverage of linear configurations are rather better, but at the expense of large array rotations and different size telescopes. The need for OPD compensation is also a serious drawback. Among linear configurations, the 4-telescope regular array (modulation between two regular DACs) seems very attractive.

4.4 Inherent internal modulation

A new kind of internal modulation has recently been proposed by Karlsson (personal communication). The principle of inherent internal modulation is presented hereafter, followed by Karlsson's example (referred to as the "Nils" configuration) and some personal investigations.

4.4.1 Principle

The goal of inherent modulation is to perform internal modulation without extra $\pi/2$ phase shift between the outputs of the two sub-nulling interferometers. This can be done if these outputs already have the properties of the S_{12} and S_{21} signals which are detected with internal modulation, i.e., the transmission maps $R_1(\vec{\theta})$ and $R_2(\vec{\theta})$ of the two sub-interferometers must be asymmetric, and conjugated by central symmetry. In order to achieve non symmetrical transmission maps, the two sub-interferometers must have a complex entrance pupil, as proven in appendix A. The central symmetry relation between the two transmission maps is easily obtained by inverting the signs of the applied phase shifts (so that $R_1(\vec{\theta}) = R_2(-\vec{\theta})$, see equation (3.4)).

We have seen in section 3.2.3 how five telescopes on a circle with a complex entrance pupil can give a transmission map without central symmetry, with a θ^{-4} rejection rate. Moreover, as proven in appendix B.1.3, there is a double infinity of complex configurations with five telescopes on a circle yielding a θ^{-4} rejection rate when the positions of the telescopes are fixed. The following example (Nils configuration) uses two "inverted" sets of phase shifts for five telescopes on a regular pentagon, in order to produce the two required outputs.

4.4.2 The Nils configuration

The geometry of the Nils configuration and its two sub-nulling interferometers is illustrated in figure 4.30. The five telescopes are arranged in a regular pentagon. The two sub-nulling interferometers are simply two "inverted" replica of the 5-telescope circular array described in section 3.2.3, which was first proposed by Léger *et al.* [31]. The two mirrored transmission maps R_1 and R_2 of the sub-interferometers are presented in figure 4.31.

As in the case of internal modulation, the modulation efficiency is defined as $\frac{|S_1 - S_2|}{2}$ where S_1 and S_2 are the output signals of the two sub-interferometers. The following

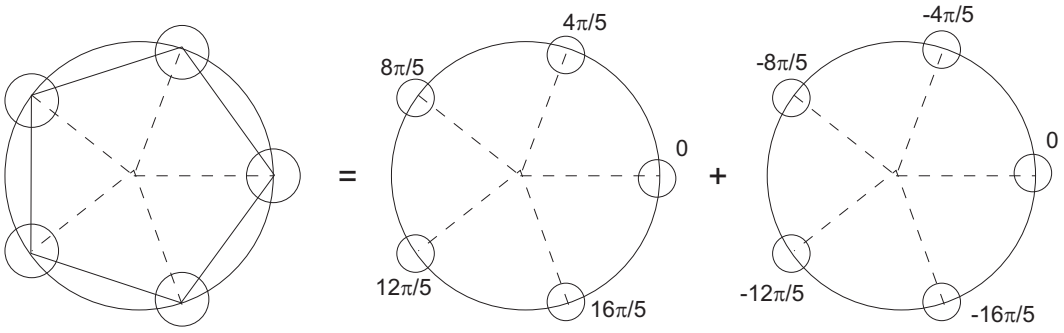


Figure 4.30: The Nils interferometer with its two sub-nulling arrays. The applied phase shifts are indicated near each telescope for each sub-interferometer.

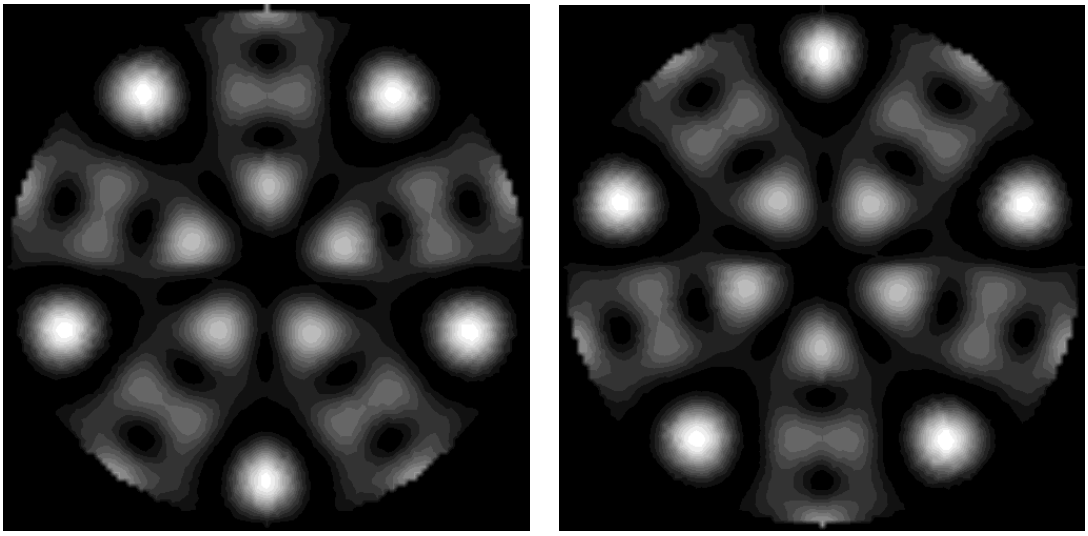


Figure 4.31: Transmission maps for the two sub-nulling interferometers of the Nils configuration. Field of view is 500×500 mas, array radius 15 m and observation wavelength 10 μm as usual. These maps are conjugated by central symmetry, which allows internal modulation between them.

modulation efficiency was found with an IDL code:

- Mean modulation efficiency: 0.060
- Maximum modulation efficiency: 0.298
- RMS of modulation efficiency: 0.061

These are indeed exactly the same values as for the "Daisy" regular pentagon configuration (see section 4.1.2). Moreover, the modulation maps of Daisy and Nils are identical (see figure 4.32). We have already noticed in section 4.1.2 that nulling interferometers with the same geometrical arrangement yield the same modulation map when internal modulation

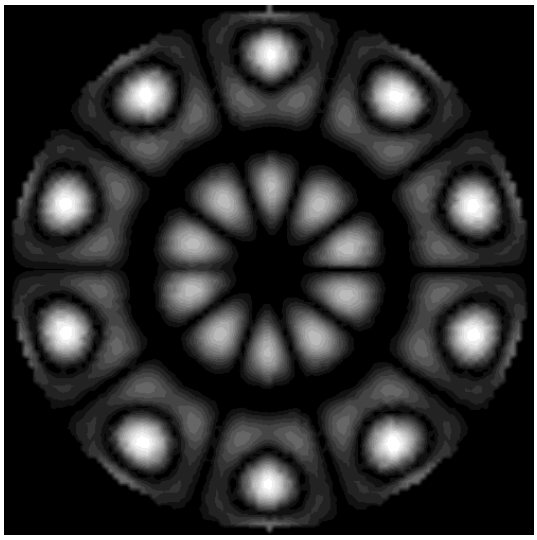


Figure 4.32: Modulation efficiency of the Nils configuration. A uniform modulation crown can be obtained by a single 18° rotation (see figure 4.7)

is used, whatever the recombination scheme. This property seems to remain true when inherent internal modulation is used. So, inherent modulation should not give better modulation efficiencies than classical internal modulation.

However a new way of using the output signals S_1 and S_2 of the two sub-nulling interferometers can be designed: instead of continuously detecting the two signals on two detectors (as was proposed at the end of section 3.2.5 to optimize the overall SNR), one could redirect all the collected light alternately to a single output (S_1 or S_2) and alternately detect these "full light" outputs on a single detector. This can be done by placing an optical switch behind each telescope in order to choose toward which output the beams will be directed. It can easily be shown that this does not change the final SNR. The only difference is that only one detector is required with inherent modulation. This is quite a good thing since detector cooling needs cryogenic technology, which consumes a lot of power and sets the maximum lifetime of the mission due to the limited cryogenic fluid reserves.

Note that the use of a single detector is also possible in the case of classical internal modulation, but at the expense of the SNR: since the two outputs S_{12} and S_{21} are simultaneously formed by a single beam-splitter, it is not possible to redirect all the light to only one output. So, when alternate detection on a single detector is chosen, there is always one output that is not detected, and so one half of the signal is lost.

4.4.3 A new 5-telescope recombination scheme

We have already seen several times that configurations involving sub-interferometers composed of five telescopes badly suffer from the recombination losses associated to third order recombination schemes: the recombination efficiency of each sub-interferometer is only 5/8.

This is the reason why the overall modulation efficiency is significantly lower than for other configurations. In order to improve this efficiency, a new recombination scheme is proposed hereafter (figure 4.33), for which the recombination efficiency increases to 5/6, so that the modulation efficiency is multiplied by a factor 8/6. This scheme uses $\frac{1}{3}$:$\frac{2}{3}$ beam-splitters at each telescope output, exploiting the fact that, for each sub-interferometer, only two of the five beams cross three beam-combiners whereas the other three beams cross only two beam-combiners (and experience therefore less losses). Unfortunately, all the collected light cannot be directed to a single output any more, so that two detectors must be used in order to get an optimal SNR.

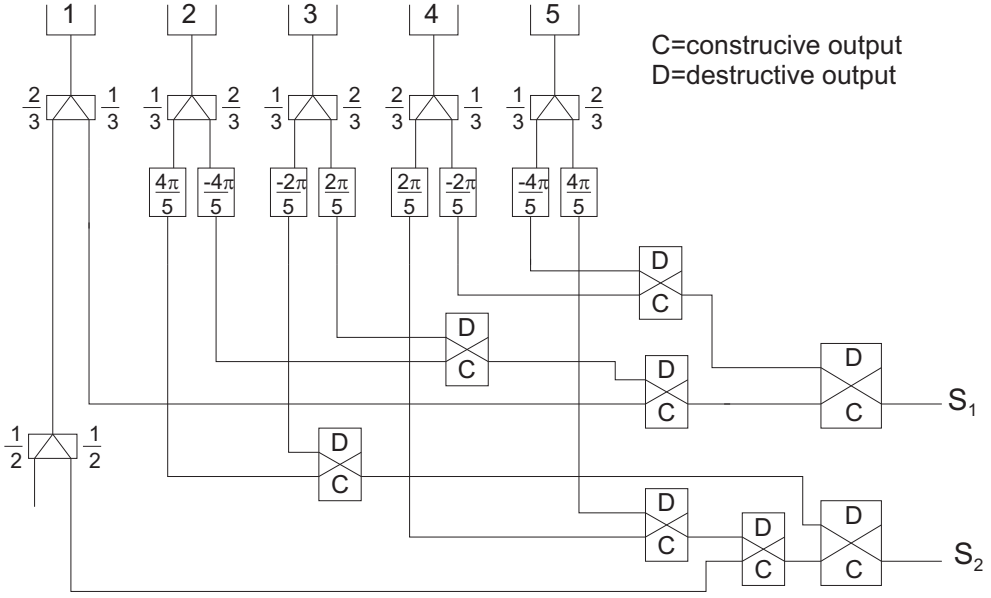


Figure 4.33: Recombination scheme for the Nils configuration. The losses are reduced by introducing $\frac{1}{3}$:$\frac{2}{3}$ beam-splitters at each telescope output.

With this new recombination scheme, the modulation efficiency increases to:

- Mean modulation efficiency: 0.080
- Maximum modulation efficiency: 0.397
- RMS of modulation efficiency: 0.081

Note that this recombination scheme could be adapted to any of the interferometers performing internal modulation between two FTRCs (section 4.1.2), multiplying the modulation efficiency by 4/3. But even with this trick, the maximum modulation efficiency hardly reaches 40%.

4.4.4 Other configurations with inherent modulation

Other configurations with five telescopes on a circle can easily be found, since for each set of angular positions of the telescopes, a double infinity of complex configurations with a

θ^{-4} rejection rate can be found. But the efficiency of all those configurations is tampered by the third order recombination scheme. Configurations with a maximum efficiency above 40% should not exist, even if the new recombination scheme is used (remind that the higher maximum efficiency was 30% in section 4.1.2).

In order to improve the recombination efficiency, let us investigate inherent modulation between four telescopes, for which a second order recombination scheme is sufficient. It is proven in appendix B.1.2 that complex configurations with four telescopes on a circle cannot give a θ^{-4} rejection rate, so that asymmetric transmission maps cannot be obtained with circular 4-telescope configurations. On the other hand, appendix C.2 shows that, for any given linear geometry, there exists a double infinity of complex configurations with a θ^{-4} rejection rate, depending on the relative values of the phase shifts ϕ_3 and ϕ_4 with respect to the $\phi_2 = 0$ reference. Inherent modulation can therefore be implemented with four in-line telescopes.

Linear configurations with four telescopes

We have seen in section 4.3 that delay compensation is a complicated task with linear configurations: simple solutions have only been found for symmetrical arrays. We will therefore restrict our study to symmetrical configurations with respect to the center of the array, for which symmetrical telescopes (1-4 and 2-3) have the same size ("A-B-B-A" configurations). A systematic study of these configurations has yield some interesting properties:

- The outer telescopes must be smaller than the inner telescopes for a symmetrical array to exist (diameter ratio B/A > 1).
- For each diameter ratio, a complex configuration can be found whatever the (symmetrical) geometrical configuration.
- For each diameter ratio, the maximum modulation efficiency happens when the distances between the four telescopes are of the same order. Indeed when two telescopes are close together, the interferometer becomes nearly equivalent to a Bracewell (when the inner telescopes are close to the outer ones) or to a DAC (when the inner telescopes merge), so that the entrance pupil becomes nearly real, and the asymmetry is broken.

The search for an optimal modulation efficiency has led to the following configuration:

	Diameter (arbitrary units)	**Position** (arbitrary units)	**Phase shift** (rad)
Telescope 1	$D_1 = 0.62$	$L_1 = 0$	$\phi_1 = 1.793$
Telescope 2	$D_2 = 1$	$L_2 = 0.55$	$\phi_2 = 0$
Telescope 3	$D_3 = 1$	$L_3 = 1$	$\phi_3 = 4.367$
Telescope 4	$D_4 = 0.62$	$L_4 = 1.55$	$\phi_4 = 2.576$

The equalization of the optical paths can be achieved as in figure 4.29. The associated modulation efficiency is of the same order as for linear arrays with internal modulation:

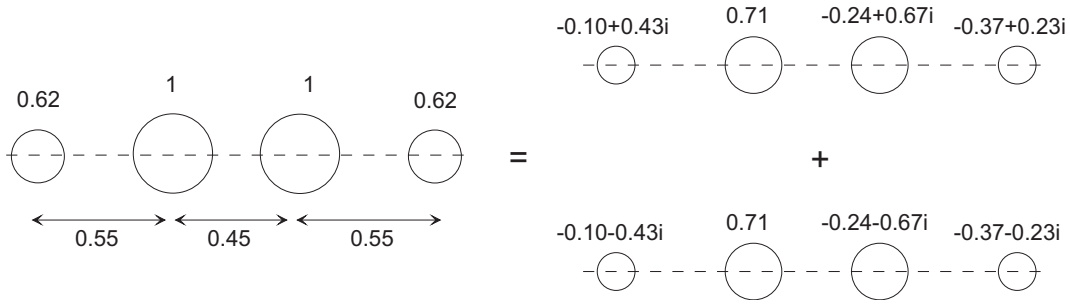


Figure 4.34: The optimal linear interferometer with its two sub-nulling arrays. The complex entrance pupil coefficients are indicated near each telescope for each sub-interferometer.

- Mean modulation efficiency: 0.163
- Maximum modulation efficiency: 0.473
- RMS of modulation efficiency: 0.166

The asymmetric outputs S_1 and S_2 of the two sub-interferometers are shown in figure 4.35, and the modulation map in figure 4.36 together with the superposition of three rotated maps. The modulation map shows two bright vertical lines on each side of the central null. When rotated, this map produces an almost uniform modulation crown, about 50 mas wide, as was the case for the linear configurations with internal modulation. Note that the modulation efficiency of these new configurations is not as high as when two DACs were recombined in a 4-telescope array. This is not a contradiction to the "same geometry - same efficiency" property since the sub-nulling interferometers do not have the same number of apertures this time.

Non linear configurations with four telescopes

In the course of the last two chapters, we have seen that Darwin's interferometric array needs at least four telescopes. But it was proven that circular configurations with four telescopes cannot achieve the required chopped output since one and only one GAC exists for a given circular configuration. So with four telescopes, only linear configurations have been considered. Those configurations have good modulation efficiencies, but a line-shaped modulation map, which is not desired. It would therefore be worth to see if an arbitrary complex configuration with four telescopes, not restricted to lie on a circle any more, could achieve both a good modulation efficiency and a more uniform transmission map, by using inherent modulation. This kind of configuration has not been considered yet because the optical path delays cannot be easily compensated in this case, since no point is equidistant from the four telescopes and since long delay lines must be avoided. The only way to compensate the optical path delays would be to add a constant optical path to one (or more) interferometer arm, but the scale of the array could not be changed any more.

The analytical developments for arbitrary 4-telescope arrays are presented in appendix D. The only difference with circular configurations is that the lengths L_k of the interferometer

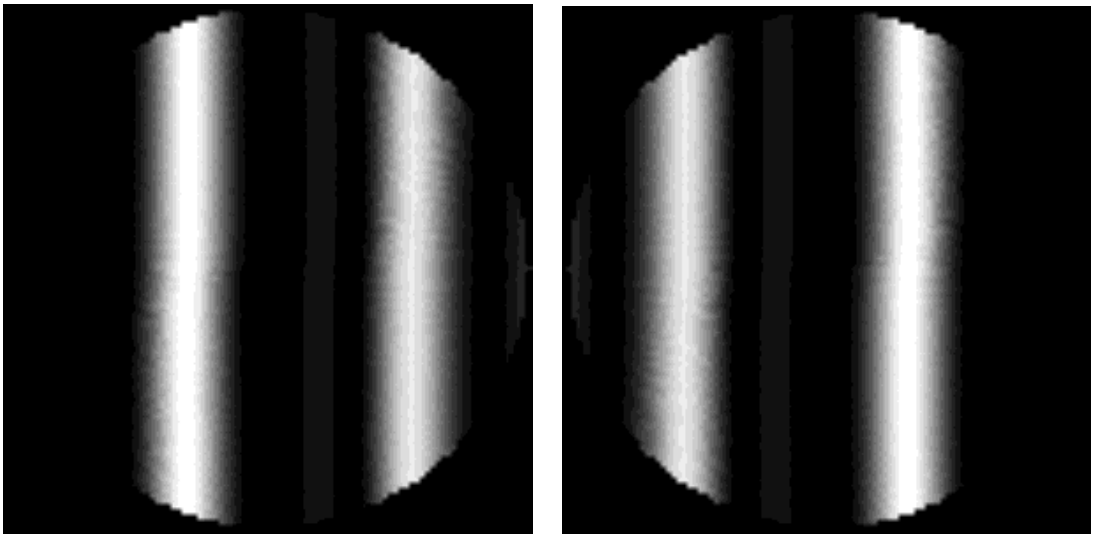


Figure 4.35: Transmission maps for the two sub-nulling interferometers of the optimal linear configuration. Field of view is 500×500 mas, array length 50 m and observation wavelength 10 μm as usual. These maps are conjugated by central symmetry, which allows internal modulation between them.

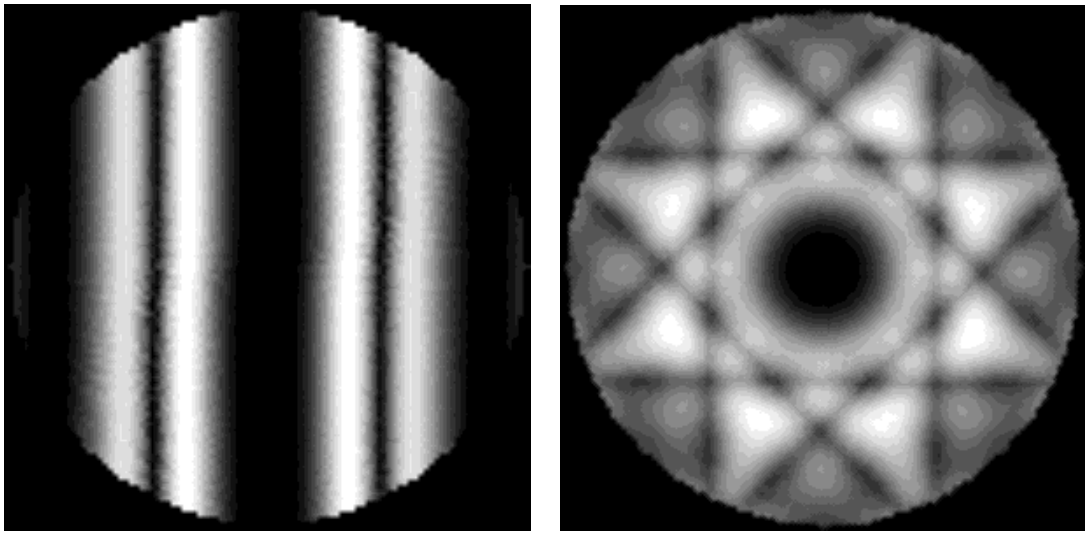


Figure 4.36: Modulation efficiency map for the optimal linear configuration. The right-hand map is deduced by superimposing three maps shifted by 60° with respect to each other.

arms are not equal any more. As usual, we take telescope 1 as a reference, such that $L_1 = 1$, $D_1 = 1$, $\delta_1 = 0$ and $\phi_1 = 0$. It is proven in appendix D that six free parameters must be fixed in order to define a single configuration. These parameters are the angular positions δ_2, δ_3, δ_4 of the three remaining telescopes, and the characteristics of telescope 2: L_2, D_2 and ϕ_2.

Arbitrary 4-telescope configurations with inherent modulation have not been fully studied, because of the large number of free parameters involved. The following approach has been used:

1. Take L_2 as a fixed parameter, and choose three arbitrary angles δ_2, δ_3 and δ_4.
2. For these four fixed parameters, find the parameters D_2 and ϕ_2 which give the best modulation efficiency.
3. See if the modulation efficiency can be improved by changing the angular parameters, and try to converge to a maximum efficiency by varying those parameters. The optimal values of D_2 and ϕ_2 do not change much during this phase.
4. The configuration resulting from this procedure is a local maximum of the modulation efficiency.

By repeating this procedure for different sets of initial parameters L_2 and δ_k, it was discovered that the arbitrary configurations always converge towards linear configurations in order to increase their modulation efficiency. Three examples of local maxima are presented below.

Local maxima of modulation														
	Optimum parameters						Deduced parameters						Modulation	
	L_2	δ_2	δ_3	δ_4	D_2	ϕ_2	L_3	L_4	D_3	D_4	ϕ_3	ϕ_4	Max	Mean
1	1.5	50°	128°	250°	1	1.88	-1.69	-2.69	0.64	0.63	-2.58	-1.81	0.47	0.17
2	2.5	27°	161°	307°	0.3	2.61	-0.77	0.68	0.95	0.81	-1.98	2.04	0.44	0.14
3	1.2	52°	208°	300°	0.88	2.07	-0.99	3.84	0.96	0.20	-2.00	2.63	0.42	0.16

Figure 4.37 shows that these three configurations are in fact linear configurations. This suggests that the best modulation efficiencies are reached for linear arrays, as far as inherent modulation between four telescopes is considered. Moreover symmetrical configurations (as configuration 1 in figure 4.37), which are the only ones to allow simple delay compensation, seem to have the highest efficiencies among linear configurations. This last remark should be taken with care since a complete study has not been done.

4.4.5 Conclusion

Inherent modulation is quite a promising technique, especially in the case of four telescopes in a symmetric linear configuration. In that case, a maximum efficiency of almost 50% can be reached, and only one detector has to be used. But the drawbacks associated to linear configurations remain (see section 4.3.5).

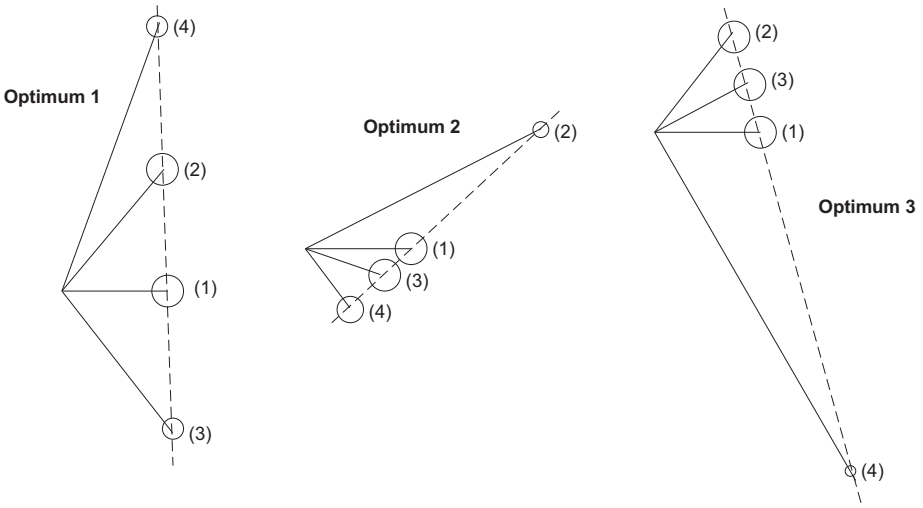


Figure 4.37: These three configurations have a local maximum in the modulation efficiency, in the general case of arbitrary 4-telescope configurations. All three are in fact linear configurations.

4.5 New kinds of configurations

4.5.1 A dual 3-telescope nulling interferometer

A new kind of nulling interferometer, which provides a direct measurement of both visibility amplitude and phase, has recently been proposed by Velusamy *et al.* [55]. Six telescopes are configured to form two identical DACs, with baseline $b \sim 30$ - 50 m (see figure 4.38). They are placed on a longer baseline B ($\sim$ 100 - 1000 m), and combined with a variable phase shift in order to produce a chopped output. A short description of this interferometer and of its capabilities is presented hereafter.

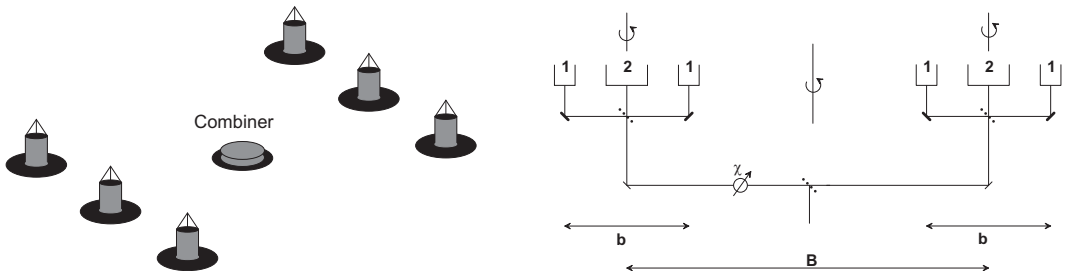


Figure 4.38: An artist sketch of the dual 3-telescope nulling interferometer, and a schematic of the beam-combination. A chopped output is obtained by switching a variable phase shift χ in one of the nulled beams.

The basic design is that of a phase shifting two-element interferometer which measures both the sine and cosine fringe amplitude. The sine and cosine fringes on the planets are

obtained with the long imaging baseline B between the nulling elements. Because nulling and imaging are independent, the characteristics of the null response (determined by b) and spatial resolution (determined by B) are totally de-coupled. This is a big advantage because in other configurations, increasing the baseline to achieve higher spatial resolution would lead to narrower null which will result in resolving the central star.

If R_{DAC} is the response of a DAC, then the response of the interferometer formed by the nulled beams with a relative phase shift χ reads

$$I_{out} = R_{\mathrm{DAC}}(1 + \cos(\Phi - \chi)) \tag{4.5}$$

where $\Phi = 2\pi\vec{B} \cdot \vec{S}/\lambda$ is the external optical path difference between the two DACs, if $\vec{S}$ represents the direction of the target planet. The two chopped outputs are obtained by switching the phase shift between 0° and 180° (cosine chop) or between 90° and 270° (sine chop):

$$I_{Cos-chop} = I_{out}(\chi = 0) - I_{out}(\chi = \pi) = 2R_{\mathrm{DAC}} \cos\Phi \ , \tag{4.6}$$

$$I_{Sin-chop} = I_{out}(\chi = \pi/2) - I_{out}(\chi = 3\pi/2) = 2R_{\mathrm{DAC}} \sin\Phi \ . \tag{4.7}$$

A uniform nulling map is obtained by rotating the two DACs, and the (u, v) plane is sampled by rotating the global interferometer around the central hub. Since $\Phi = 2\pi\vec{B} \cdot \vec{S}/\lambda = 2\pi(ux + vy)$, the chopped outputs (4.6) and (4.7) measure the real and imaginary parts of the complex visibility at the $u - v$ coordinates corresponding to $\vec{B}$. The observed complex visibility reads

$$V_{obs}(u, v) = \int_x \int_y R_{\mathrm{DAC}}(x, y, \beta) I(x, y) \exp(2\pi i(ux + vy))\, dx\, dy \ , \tag{4.8}$$

where β is the position angle of the baselines of the DACs (which are supposed to remain parallel). This equation allows to reconstruct the image of extra-solar planets by a simple Fourier transform of the observed visibilities. The observed intensity distribution is the true image attenuated by the null response: $I_{obs}(x, y) = R_{\mathrm{DAC}}(x, y, \beta) I(x, y)$.

4.5.2 Internal modulation between two coronagraphs

A strong limitation to single aperture coronagraphs is their low resolution. The use of two separate coronagraphic telescopes could both improve their resolution and eliminate the background signals by means of internal modulation. Two coronagraphs can be seen as two sub-nulling interferometers. If their outputs are recombined with a $\pm\pi/2$ phase shift, the two resulting signals S_{12} and S_{21} would simply have two shifted fringe pattern consisting in equidistant vertical lines. Internal modulation between these two outputs would improve the resolution since the distance between two successive fringes is inversely proportional to the distance between the two telescopes, and would also eliminate the background signals which have the same contribution in S_{12} and S_{21}. This new possibility has the unique advantage to only require two telescopes. Two detectors would be needed in order to continuously detect the S_{12} and S_{21} signals.

The new generation of coronagraphs, such as Roddier's phase-mask coronagraph or Baudoz' hybrid coronagraph (respectively defined in sections 2.3.3 and 2.3.4), would probably be well suited for this new kind of configurations. On the other hand, the Achromatic Interfero Coronagraph cannot be used as such because it produces two symmetrical images of an off-axis planet, which would thus be cancelled with internal modulation (just as the symmetrical exo-zodiacal cloud).

4.6 Configurations without internal modulation

We have seen in chapter 3 that the first proposed configurations for the Darwin and TPF missions did not use internal modulation. The planet signal modulation was achieved by rotating the interferometer array. Since 1997, such configurations are not considered any more. Let us briefly summarize the reasons why external modulation has been abandoned.

- *Modulation frequency.* In order to reduce the background drifts and other $1/f^{\alpha}$ type noises, a modulation at "high" frequency is needed. This cannot be achieved with external modulation since the rotation frequency of a 50 m-long interferometer hardly exceeds 10^{-4} Hz.

- *Continuous rotation.* A lot of time is spent in rotating the array, which consequently reduces the observation time.

- *Number of telescopes.* In order to get rid of the spurious exo-zodiacal signal, an asymmetric transmission map is required. This implies that at least four telescopes are used, or even five if a circular configuration is desired. These are the same numbers as when internal modulation is used.

- *Modulation efficiency.* The theoretical modulation efficiency of the planet signal is determined by the recombination efficiency of the interferometer[9]: the factor affecting the planet signal varies between zero and the maximum recombination efficiency as the planet crosses bright and dark fringes. But remember that only the odd multiples of the fundamental frequency are used, since the even frequencies also include a strong exo-zodiacal contribution (see section 3.2.3). All in all, the modulation efficiency should not be better than when internal modulation is used.

4.7 Conclusion: a menagerie of configurations

The systematic study of the Darwin aperture configurations undertaken in this chapter has led to a huge number of configurations with a large variety of shapes and efficiencies. Now the problem we are faced with is the choice among this menagerie of configurations. This important topic will not be comprehensively discussed here. We only present some criteria that might lead to an optimal configuration.

[9]The recombination efficiency reaches 100% for a 4-telescope interferometer, but only 60% for a 5-telescope one.

4.7.1 Criteria for an optimal configuration

- *Modulation efficiency.* This is probably the most important criterion. The maximum efficiency (78%) happens for the 5-telescope linear configuration (5Lin2DAC) with internal modulation between two regular DACs. Among circular configurations, the best efficiency (65%) happens for the 6-telescope configuration with internal modulation between two GACs (RL2(1,1,0.707,0,0,0.707)). Note that the 4-telescope linear configuration with equal size telescopes and modulation between two irregular DACs (4LinEqual2DAC) is also very interesting (maximum efficiency 67%). The configurations with a third order recombination scheme have lower efficiencies.

- *Uniformity of the modulation map.* This is an important criterion in order to achieve planet detection with a minimum time and power consumption. The 1-2-3-5 configuration with an angular parameter of 40° (RL3(1,0.449,0,0.293,0,0.844)) is the best on that point (max 41%), together with the Daisy (or Nils) configuration (max 30% or 40%). On the other hand, large rotations must be applied to linear configurations in order to make their modulation map uniform. Moreover, the first maxima of linear configurations are sometimes faint and narrow.

- *Number of detectors.* On that point, configurations with inherent modulation are the best since they only use a single detector. Other configurations generally use two detectors, except in the case of three sub-interferometers, where six detectors must be used. This is a serious drawback to the regular hexagon and other configurations of that family.

- *Number of telescopes.* Linear configurations typically use only four telescopes, whereas circular configurations need at least five telescopes. This is not a decisive advantage. A maximum of six telescopes is a practical limitation due to the size of an Ariane 5 fairing. Note that the use of two coronagraphic telescopes drastically reduces that number.

- *Equal size telescopes.* It is preferable that all telescopes have the same size in order to reduce the manufacturing cost. Circular configurations were designed to account for this fact, but on the other hand, only one linear configuration (the 4LinEqual2DAC) achieves the size equality.

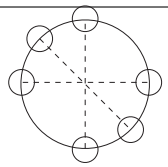
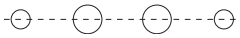
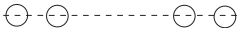
- *Recombination optics.* On that point, the linear configurations have the serious drawback that the light beams must be reflected on neighboring telescopes in order to achieve the optical path equality. In the particular case of the 5Lin2DAC configuration, a simple way to compensate the optical delays has not even been found. On the contrary, circular configurations can be quite simple, especially if only two GACs are involved (as in the RL2(1,1,0.707,0,0,0.707) configuration).

- *Versatility.* The regular OASES configuration is very versatile since more than one recombination scheme can be used with the same geometry, resulting in different efficiencies and rejection rates. The RL3(1, 0.449,0,0.293,0,0.844) and RL3(1,1,0,0.532, 0.532,0) are two circular configurations (composed of two triangles shifted by 40°)

which show a good complementarity: exceptional sky coverage for the first, very high efficiency (59%) for the second.

4.7.2 Final configuration choice

The final choice for a configuration will have to take into account all these criteria, and probably others as well. Based on the previous criteria, and on my personal opinion, the best configurations are the RL2(1,1,0.707,0,0,0.707) among circular configurations, and the 4Lin2DAC or 4LinEqual2DAC among linear configurations. Their characteristics are summarized in the table below.

			
	RL2(1,1,0.7,0,0,0.7)	4Lin2DAC	4LinEqual2DAC
Max efficiency	65%	51%	67%
Mean efficiency	13%	18%	18%
Map uniformity	Fair	Fair	Bad
Detectors	2	2	2
Telescopes	6 (equal size)	4 (diff. size)	4 (equal size)
Recombination	Very easy	Complicated (refl.)	Complicated (refl.)
Versatility	Bad	Good	Fair

Finally a very important criterion which is not discussed in this work is the ability to reconstruct an image of the planet from the nulled signal. Except for a cross-correlation method designed by Angel and Woolf [3], very little work has been done on that topic yet. Modelization of the influence of internal modulation on imaging will be an important step for the Darwin mission, and will probably shed a new light on the final configuration choice.

Chapter 5

Nulling with the VLTI

5.1 Context and scientific goals

We have proven in section 3.4.2 that the exo-zodiacal light is one of the main contributors to the background noise for the DARWIN mission. Its level was computed assuming the same amount of dust as in our Solar System, and its total contribution was found to be about 5 times smaller than the contribution from the local zodiacal light (for a spacecraft at 1 AU from the Sun). But the assumption of a "Solar System twin" is not based on any physical argument, and DARWIN's signal-to-noise ratio (SNR) could well be limited by the exo-zodiacal light if the dust cloud was much denser than ours. An exo-zodiacal cloud 10 times brighter than our own would only reduce the SNR by a factor 1.5 (or equivalently, increase the integration time by a factor 2.2), but an exo-zodiacal cloud 100 times brighter than ours would increase the integration time by a factor about 14, which is not acceptable. Thus before precisely defining the DARWIN (or TPF) mission, direct measurements of the amount of warm dust around nearby stars should be undertaken. This is one of the main conclusions of the "Exo-zodiacal Dust Workshop" that was held in October 1997 at the NASA Ames Research Center[1].

Exo-zodiacal dust surveys can be carried out with infrared facilities such as SIRTF, SOFIA or ALMA if available early enough, or even with the HST[2], but will mostly rely on ground-based infrared interferometers such as the Keck Interferometer, the Large Binocular Telescope (LBT) and the Very Large Telescope Interferometer (VLTI). In order to detect exo-zodiacal clouds against the blinding star, these interferometers must use the nulling technique. While a nulling instrument is already scheduled for both Keck (2001) and LBT (2004), a concrete project is not planned yet for the VLTI. An agreement between ESA and ESO has recently been found concerning the installation of a nuller as a guest instrument on the VLTI, but its characteristics are still to be defined. The nuller should achieve two parallel goals:

1. Validation of the nulling technique.

[1]Backman's report available at http://astrobiology.arc.nasa.gov/workshops/1997/zodiac/index.html.

[2]NASA is now studying a coronagraphic capability as part of an imager (WF3) that could be installed on HST in 2002. This instrument should have the capability to detect 100-zodi clouds and resolve their orientation around nearby stars.

2. Characterization of exo-zodiacal clouds.

In order to be beneficial to the Darwin mission, the nulling instrument should be sensitive to exo-zodiacal light down to the level which tampers Darwin's performances, i.e., exo-zodiacal clouds 10 times brighter than our own. If a large proportion of stars are found to have a zodiacal cloud brighter than this limit, the size of the Darwin apertures should probably be reconsidered.

5.2 Exo-zodiacal cloud properties

Before discussing the nulling potential of the VLTI, let us first briefly discuss the structure and properties of exo-zodiacal clouds. This section is mostly based on the Executive Summary of the Exo-zodiacal Dust Workshop and on Backman's report [4], both available at the URL mentioned in a previous footnote.

5.2.1 Dust sources

The dust in our solar system has three principal constituents: asteroidal and cometary debris and interstellar grains (see TPF [9], chapter 5). The main asteroid belt seems to be the main source for the smooth zodiacal cloud: debris are released by asteroid collisions and erosion of small asteroids. Its relative contribution is supposed to lie between 50% and 90%. Symmetric bands of thermal emission near the ecliptic have been identified as "families" of asteroid collision debris that drift past the Earth toward the Sun from sources in the main asteroid belt. Further out, the Kuiper belt can supply grains to the inner Solar System via radiative drag due to sunlight, known as the Poynting-Robertson (PR) effect.

COBE data indicate the presence of some local dust that is not distributed following the wedge of the main belt asteroid debris and is thus probably cometary in origin. Extensive dust trails fill the orbits of periodic comets with mm-sized dust and meteors, which are eventually perturbed into the zodiacal dust cloud and sporadic meteor population. Comet activity and collisions should contribute between 10% and 50% to the zodiacal dust near Earth.

The last contributor is the interstellar medium, which is expected to be a significant source of the smallest particles in the outer solar system (cold dust). Its contribution to warm dust near the Earth should however be small (about 0.1%). Note that warm dust (temperature about 300 K) is the most important background source for Darwin because its flux peaks around 10 μm, and because it is mostly concentrated between 0.01 AU[3] and 2 AU (thus covering the habitable zone).

5.2.2 Dust density

One of the most important conclusions of the Exo-zodiacal Dust Workshop is that we do not have the understanding at present to allow theoretical prediction of the amount of dust

[3]The dust destruction temperature is about 1500 K, which corresponds to a sublimation radius of about 0.01 AU.

at terrestrial temperature in a given planetary system even if it is known to contain a cold dust disk and/or planets. One can however define approximate bounds to the amount of dust by means of theoretical and observational considerations:

- *Lower limit.* The necessary presence of comets in exo-systems might imply a very uncertain general lower limit of 1/10 the zodiacal dust density. Planetesimal models of planet formation produce an asteroid belt in only 50% of the trials.
- *Upper limit.* IRAS measurements indicate a few cases among nearby stars of more than 100-zodi clouds of warm dust. On the other hand, approximately 15% of stars of spectral types A, F, G and K should have 100 times or more cold dust than our own solar system. Theoretical calculations suggest that there may be a natural upper limit to steady state density of warm dust, because of mutual dust collisions.

Another prediction of theoretical models is the time-dependence of the cloud density (see TPF [9]). Collisions of 10 km-radius bodies should occur in the main belt about every 10^7 years. One such collision could completely recreate the present zodiacal cloud. A model history of density shows a general decline as grain fall into the Sun due to PR drag, but punctuated by large transient increases due to the subsequent collisions between fragments. Planet searches should be directed towards older stellar systems to avoid generally higher dust density.

5.2.3 Dust distribution and brightness

IRAS data have brought considerable information on dust density in our zodiacal cloud. Based on these data, a model for the zodiacal cloud has been published. This model has already been used in section 3.4.2, equation (3.24), in order to evaluate the contribution of a "Solar-type" exo-zodiacal cloud to the total noise for the Darwin mission. It relies on a grey-body approximation, with emissivity $\varepsilon(r) = 1.114 \times 10^{-7} r^{-0.8}$ (r being the distance from the star in AU), and dust equilibrium temperature $T(r) = 266\, r^{-0.36}$.

Figure 5.1 shows the face-on optical depth of the smoothly varying portion of the Solar System zodiacal dust (as seen from 10 pc) as a function of radius from near the Sun to 100 AU for the COBE model parameters. The COBE model spatial distribution $\tau \sim 0.71 \times 10^{-7} r^{-0.39}$ was assumed valid from near the Sun to 100 AU (note that the IRAS model has a steeper gradient than the COBE model). The same figure also shows typical temperatures in the cloud assuming emissivity = 1. The corresponding face-on surface brightness is displayed on the right-hand side of figure 5.1 at wavelengths of 8, 10 and 12 μm. The hot inner regions of the disk are much brighter in the mid-infrared range because of their higher temperature and optical density.

The integrated flux from the dust sublimation radius outward is shown in figure 5.2 for the IRAS model. It is particularly interesting to note that 90% of the flux is included inside a radius of about 1 AU. This implies that, in order to characterize the exo-zodiacal IR emission, the nuller should have a spatial resolution better than 1 AU, i.e., better than 100 mas in the case of a 10 pc star. A spatial resolution of order 0.1 AU is even desirable.

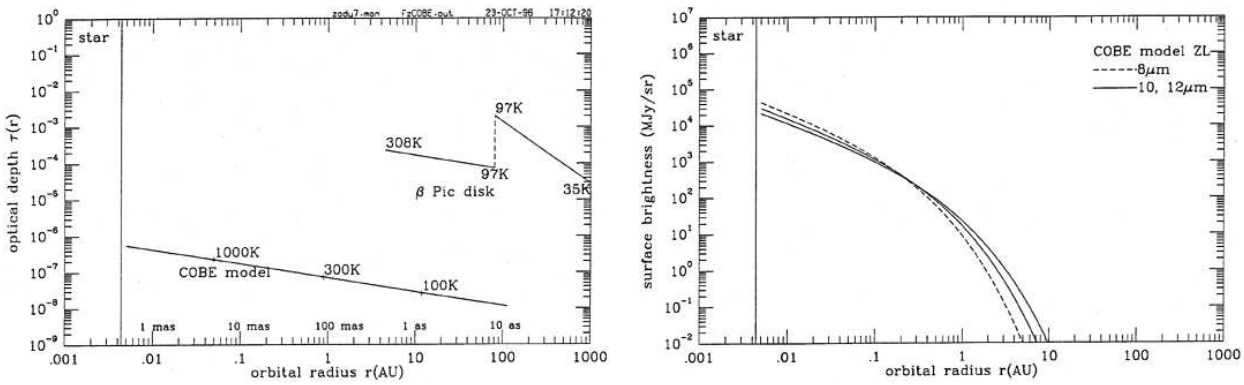


Figure 5.1: Thermal model of a 1-zodi cloud around a Solar-type star viewed from 10 pc based on the COBE zodiacal model (from Backman [4]). Left: Disk vertical optical depth and temperature versus distance from the central star, compared with a possible profile of the cold dust disk of β Pic. Right: Thermal emission surface brightness for several wavelengths around 10 μm.

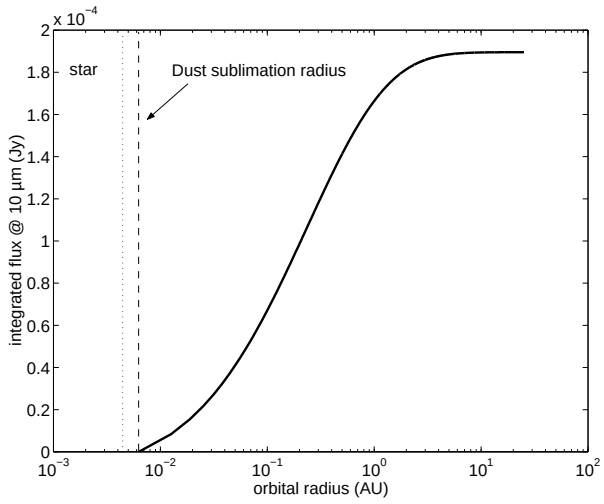


Figure 5.2: Integrated flux at 10 μm versus distance from the star for a 1-zodi cloud around a Solar-type star viewed from 10 pc, based on the IRAS zodiacal model. Note that 90% of the flux comes from the region between 0.01 and 1 AU.

5.3 Nulling with the VLTI

The ESO Very Large Telescope (VLT) at the Paranal Observatory (Atacama, Chile) is one of the world's largest and most advanced optical telescope. It presently comprises four 8.2-m reflecting Unit Telescopes (UTs) and three moving 1.8-m Auxiliary Telescopes (ATs), the light beams of which can be combined in the VLT Interferometer (VLTI).

5.3.1 Array configuration

A layout of the VLTI is shown in figure 5.3. The minimum and maximum distances between two UTs are respectively 46.6 m (between UT2 and UT3) and 130.2 m (between UT1 and UT4). The distance between two ATs on the main east-west track varies between 8 m and 128 m by steps of 8 m.

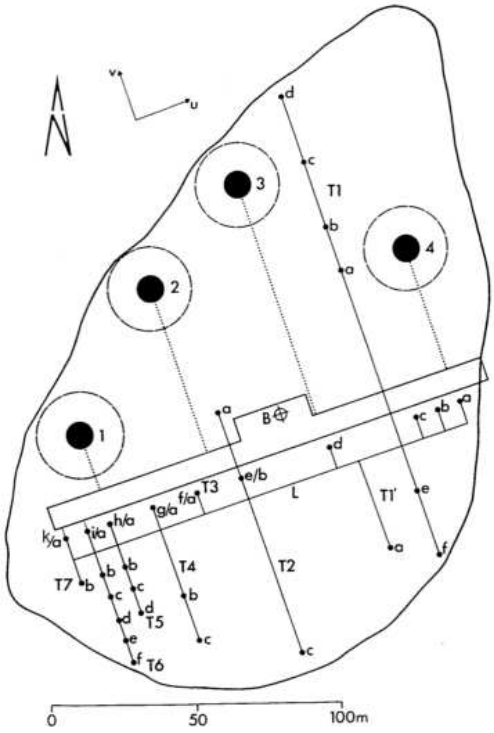


Figure 5.3: Layout of the individual VLTI elements. Unit Telescopes are shown as large, filled circles, and the possible locations of ATs indicated by small, filled circles. These circles represent the diameters of the primary mirrors when pointed at the zenith. Solid lines show the location of tracks for transporting ATs between stations.

With this configuration, there seems to be three main possibilities for a nulling instrument[4]:

- *Bracewell interferometer between two UTs.* This would be the same nulling configuration as for the Keck and the LBT, but with a different baseline. In order not to over-resolve the exo-zodiacal cloud, and by the same way, not to resolve the stellar disk, the minimum distance between two UTs should be chosen. With a 46.6 m baseline, the angular separation between the central null and the first bright fringe is 22 mas for an observation wavelength of 10 μm. This angular resolution fulfills the high resolution requirement (22 mas correspond to 0.22 AU for a typical system at 10 pc).

[4]The following discussion is largely based on two "brainstorming" meetings that were held at the Meudon Observatory in January 2001. I would like to thank all the participants for their enthusiasm and for welcoming me in Meudon.

- *DAC interferometer between three ATs.* In order to have a broader null, three ATs can be used in a DAC configuration (see section 3.2.2 for the definition of a DAC). In this configuration, the diameter ratio is $1{:}\sqrt{2}{:}1$, so that the effective size of the two outer telescopes is 1.27 m instead of 1.8 m. Since it uses smaller telescopes, this configuration has a lower sensitivity. The overall baseline can vary between 16 m and 128 m on the main east-west track.

- *Single aperture coronagraph.* The last simple possibility consists in using a single UT with a coronagraph. In order to have a sufficient resolution (better than 100 mas), the observation wavelength should be smaller than 3 μm. However, this configuration would be of no help in demonstrating the nulling technique.

These three possibilities are studied in more detail in the following sections.

5.3.2 Choice of the wavelength

The choice of the observation wavelength depends on three main factors: the contrast between the star and the exo-zodiacal cloud as a function of wavelength, the atmospheric emissivity in the infrared range and the thermal emission of optical surfaces. Other important factors are the spatial resolution and the residual phase errors, which depend on the wavelength. Let us first evaluate the wavelength dependence of the exo-zodiacal flux, and compare it to the stellar flux. This will give us an idea of the rejection rate to be reached.

Based on the IRAS model, the integrated exo-zodiacal flux peaks at a wavelength of about 3 μm (see figure 5.4). The curve naturally looks like a classical Planck curve, corresponding to a mean temperature of about 1000 K. This is not surprising since the hot inner part of the cloud has a great influence on the overall emission.

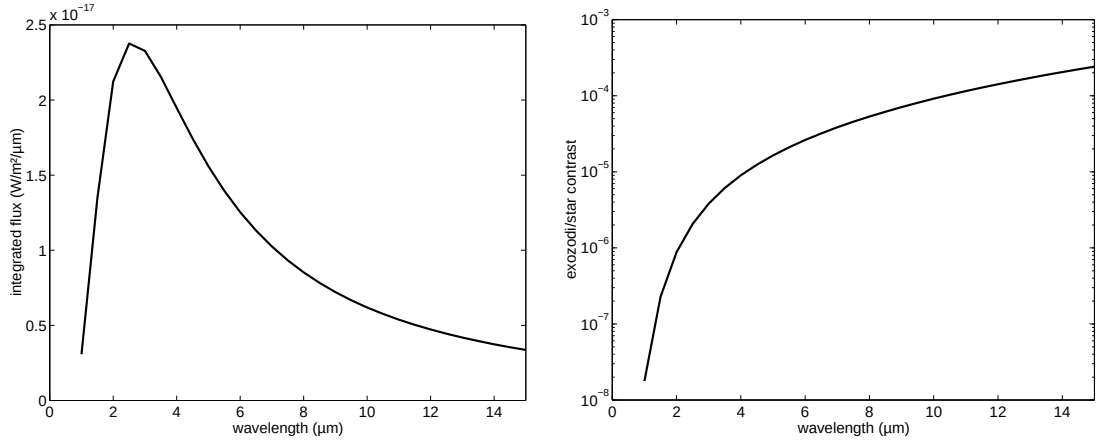


Figure 5.4: Left: Integrated flux (in $W/m^2/\mu m$) versus wavelength for a 1-zodi cloud around a Solar-type star viewed from 10 pc, based on the IRAS zodiacal model. The use of $W/m^2/\mu m$ allows an easy estimation of the total flux in a given spectral band: it is simply the area under the curve across the band. Right: Flux contrast between the exo-zodiacal cloud and the star.

The right-hand side of figure 5.4 shows the contrast between the exo-zodiacal flux and the star flux across the relevant spectral range. As in the planet/star case, the contrast is lower in the mid-infrared (about 10^{-4}) than in the near-infrared (about 10^{-6}). If the nuller performance does not allow a θ^{-4} rejection rate, the high contrast in the near-infrared range will probably not allow secure exo-zodiacal flux measurement.

Background thermal emission

Let us now evaluate the contribution from the optical surfaces and the atmosphere. The so-called "infrared atmospheric windows" are listed in the table below for wavelengths between 2 μm and 13 μm, together with the sky brightness in the four bands. The sky brightness mainly comes from the thermal emission of the atmosphere in the L, M and N bands, while the K band is dominated by OH-radical fluorescence. Note that the N band is actually divided into two parts, the larger of which lies between 10 and 12 μm. This part of the spectral band, which has a good transparency, will be used in the following discussions.

Mean sky brightness (on top of Mauna Kea, 4200 m)				
Photometric band	K	L	M	N
Mean wavelength (μm)	2.2	3.4	5.0	10.2
Band size $\Delta\lambda$ (μm)	0.40	0.55	0.3	5
Mean sky brightness (Jy/arcsec^2)	0.015	0.16	22.5	250
Sky transparency	high	fair	low	fair

Now that the possible observation wavelengths are known, the thermal emission of optical surfaces in these bands must be computed. We will see that this contribution is more than ten times stronger than the contribution from the atmosphere, because the optical surfaces of the VLTI are not cooled: the whole bunch of optical surfaces ahead the nuller is assumed to behave like a grey-body emitter of temperature 280 K, with a grey emissivity of about 60% in ideal conditions. This large emissivity comes from the fact that, before recombination, light beams undergo about 18 reflections on coated mirrors with transmission $\simeq 97\%$, yielding an overall transmission of about 40%[5]. In fact, a total emissivity of 60% is only achieved if the mirrors are perfectly clean, which is very optimistic since one cannot prevent dust particles from settling down on the mirrors. These small particles generally have a high emissivity, so that a more realistic value for the overall optical surfaces emissivity is about 100%. An estimation of their overall contribution to the thermal background is given in the table below.

Photometric band	K	L	M	N
Optical surfaces brightness (Jy/arcsec^2)	0.006	6.33	253	6581
Mean sky brightness (Jy/arcsec^2)	0.015	0.16	22.5	250
Total background (Jy/arcsec^2)	0.021	6.49	275.5	6831

The thermal background is very bright in the mid-infrared (i.e., in the N band). This is an important drawback of mid-infrared ground-based observation. Large apertures are

[5]Values from Vincent Coudé du Foresto, personal communication.

required in order to reach a good signal-to-noise ratio in an acceptable integration time. The nuller will have to get rid of the background emission by means of chopping whatever the wavelength.

The two noise sources (thermal background and star) have opposite influence on the choice of the wavelength. The final choice will therefore depend on the nulling configuration. In the next paragraphs, "back-of-envelope" signal-to-noise calculations are carried out for the three configurations. Note that the relevant wavelength for DARWIN is around 10 μm, so that the N band seems a priori the most appropriate. However, exo-zodi measurements in the K band can be extrapolated to the mid-infrared range by means of theoretical models (the IRAS model for example), if the near-infrared dust albedo is known (estimations are already available, see Backman [4]).

5.3.3 Bracewell interferometer with two UTs

In the following sections, we assume that a means to suppress the thermal background is available (some kind of chopping, see section 5.4.1). We will only consider the two most transparent spectral bands, i.e., K (λ = 2 - 2.4 μm) and N (second part of the band: λ = 10 - 12 μm), with a preference for the N band because it better corresponds to the DARWIN spectral range. In this section, only theoretical performances are evaluated, on a 10 pc Solar System twin, in order to assess the feasibility of each configuration. The actual performances have to take technical limitations into account, which are briefly discussed in section 5.4.

The Airy radius for a 8 m-class telescope working at 11 μm is about 300 mas, which corresponds to 3 AU from the central star. Since the exo-zodiacal flux is mostly concentrated in a 1 AU radius, we will reduce the field of view (and by the way reduce the thermal background) by only keeping a coherence étendue $S\Omega = \lambda^2$ at each telescope output[6]. The field of view is now 176 mas in radius. Let us first compute the photon noise from the background and compare it to the exo-zodiacal signal, which is reduced by the fringe pattern by a factor $\simeq 3$. The total signal from the thermal background is $F_{\nu,\mathrm{back}} = 6831\,\Omega = 664.8$ Jy, or equivalently 2.31×10^{10} ph-e$^-$/s over the 10-12 μm wavelength band, taking into account the optical transmission of 40% and assuming a quantum efficiency of 60% for the detector. The associated noise scales as the square root of the mean photon flux: $N_{\mathrm{back}} \sim 1.52 \times 10^5$ ph-e$^-$/s$^{1/2}$. The flux from the exo-zodi (185 μJy) is reduced to $F_{\nu,\mathrm{exo}} = 64.4$ μJy by the fringe pattern, or equivalently $F_{\mathrm{exo}} = 2.24 \times 10^3$ ph-e$^-$/s.

These values are summarized in the table below. They prove that a 1-zodi cloud is very difficult to detect: a SNR of 5 would require about 30 hours of integration, while a SNR of 2 would still require 5 hours. On the other hand, the 10-zodi level we took as a requirement for the nuller is easily detected: a SNR of 10 can be achieved in 1.3 hours.

Signal and noise for the Bracewell interferometer		
nulled exo-zodi	nulled star	background noise
64.4 μJy $\Rightarrow 2.24 \times 10^3$ ph-e$^-$/s	306 μJy $\Rightarrow 10.6 \times 10^3$ ph-e$^-$/s	152×10^3 ph-e$^-$/s$^{1/2}$

[6] We have proven in section 3.4.2 that the SNR is maximum for an optical étendue $S\Omega = 1.096\lambda^2$.

Let us now compute the rejection rate of the Bracewell interferometer for the central star, whose angular radius is 0.47 mas (Sun-like star at 10 pc). Integration of the fringe pattern over the stellar disk yields a rejection ratio $\rho = 4.25 \times 10^3$. The flux from the star being about 1.3 Jy at 11 μm, the total flux after recombination is 306 μJy, about 5 times larger than the flux from the exo-zodi. The situation would even be worse in the K band, since the fringe pattern (and hence the null) would be 5 times narrower. This is a serious drawback of the Bracewell interferometer: even after recombination, the contribution from the star is still higher than the exo-zodiacal signal. The situation would naturally be better for brighter exo-zodi or for more distant stars, which are seen with a smaller angular diameter and thus induce less leaks.

However there seems to be a means to get rid from this unwanted contribution: by slightly shifting the fringes with OPD variations in the delay line, only the stellar flux significantly varies because the star is not perfectly centered on the dark fringe any more. On the other hand, the exo-zodiacal flux does not change significantly if the OPD variations are small enough, because of its symmetry. Since the fringe pattern is known (θ^{-2} rejection rate), the contribution from the star can be evaluated with this trick, and then removed from the signal. Note that this technique would not be possible if the contribution of the exo-zodi was 100 times smaller than the residual stellar flux because of the uncertainties in stellar mid-IR spectral energy distributions.

5.3.4 DAC interferometer with three ATs

The DAC interferometer badly suffers from the small size of the Auxiliary Telescopes (effective aperture diameter of 1.27 m). Even if we only keep a coherence étendue $S\Omega = \lambda^2$ at each telescope output, the angular radius of the field of view is still large (about 1100 mas at 11 μm) compared to the size of the exo-zodiacal cloud. The flux from the thermal background being proportional to the beam étendue, its level remains the same as in the previous case, while the signal from the exo-zodi strongly decreases, proportionally to the telescope surface S. Therefore, a 1-zodi cloud is not detectable at all: a SNR of 1 in the N band would require an integration time of more than 200 hours! Detection of 10-zodi clouds would even be difficult, since 20 hours of integration are needed in order to get a SNR of 3. The use of a DAC interferometer composed of three ATs is therefore not suitable for exo-zodi detection in the mid-infrared range, and thus we are forced to go to the near-infrared (K band).

In the K band, the thermal background is not an issue any more. Its contribution is about 3.4 mJy if we keep a coherence étendue as usual (the angular radius of the field of view is 227 mas in this case). The main trouble now comes from the high contrast between the exo-zodiacal light and the central star (about 10^{-6} in the K band). A rejection rate of order 10^6 can easily be reached with a DAC since it has a θ^{-4} rejection rate, but the broad null also cancels a large part of the inner hot zodiacal cloud, and significantly reduces its total contribution. The rejection rate strongly depends on the baseline since it determines the fringe spacing. The possible distances between two neighboring ATs are 8 m, 16 m, 32 m and 64 m. The table below shows that the contrast remains quite high whatever the baseline. The 16 m baseline has the lower contrast (about 18), but it is still 3 times higher

than for the Bracewell interferometer. Moreover, the exo-zodiacal signal has dropped to 0.08 ph-e^-/s, which is not acceptable. For all these reasons, a DAC interferometer composed of three ATs does not seem appropriate for exo-zodiacal cloud characterization.

Contrast for the four possible baselines					
	no nulling	nulling (64 m)	nulling (32 m)	nulling (16 m)	nulling (8 m)
star	29 Jy	437 μJy	27.4 μJy	1.72 μJy	0.107 μJy
exo-zodi	31 μJy	1.39 μJy	0.304 μJy	0.094 μJy	0.0045 μJy
contrast	9.35×10^5	314	90.1	18.3	24

5.3.5 Coronagraphy with a single UT

As already mentioned above, the use of a coronagraph on a Unit Telescope has two main drawbacks: the inevitable use of short wavelengths (K band or shorter) to achieve a sufficient spatial resolution, and the fact that it will not demonstrate nulling interferometry for the Darwin mission. Observing in the K band with a 8.2 m telescope yields an angular resolution of 67 mas, which is sufficient for exo-zodiacal cloud measurements, but it also gives a contrast ratio as high as 10^6, which must be compensated by the coronagraph. Even if theoretically reachable with the new generation of coronagraphs, a rejection rate of order 10^6 is still to be demonstrated on the ground. To date, the study has been restricted to the densest disk. Coronagraphic techniques will probably be an indispensable complement to nulling interferometry in the future.

5.4 Technical considerations

5.4.1 Dual nulling

In order to get rid from the unwanted thermal background, a dual-nuller configuration is proposed in [4], in the case of a Bracewell interferometer. With this configuration, each aperture is divided in the pupil plane and its light fed into nulling combiners. The effective baseline for these interferometers is approximately 4 m. The output of these "aperture nullers" is then fed into the interferometer nuller already described. A second interferometer nuller is added to use the second nulled output of each aperture nuller. Four delay lines are shown in the system (figure 5.5), one preceding each input of the aperture nullers. By adjusting the optical delay lines, all combinations of aperture nulling "on" or "off" and interferometer nulling "on" or "off" can be achieved. Switching among these configurations can be used as source chopping in order to calibrate the background (see Backman [4] for details). It could also help cancel the stellar leaks. A drawback of this approach is that a great part of the exo-zodiacal emission is cancelled by the aperture nullers because of their small baseline (and hence large fringe spacing). Other possibilities have been investigated for the Keck nulling interferometer, and could overcome this drawback[7].

[7]Personal communication from B. Mennesson.

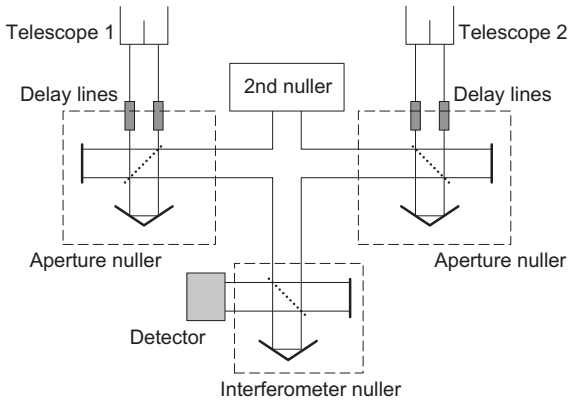


Figure 5.5: Block diagram for achromatic dual-nulling (one nuller at each aperture plus one for the 2-telescope interferometer as a whole), taken from [4].

5.4.2 Limitations to the rejection rate

The main practical limitations to the rejection rate and means to fight them are presented below (mainly based on Backman [4] and on the Meudon meetings).

- *Optical path jitter.* Deep nulls require a high degree of path length stability between the two arms of the interferometer. The rejection rate will therefore strongly depend on the performances of the fringe tracking system. A fringe sensor unit will be part of the PRIMA[8] instrument at the VLTI. The specifications for PRIMA (phase error about 30 nm rms at 10 μm) should ensure a very good null in the N band. The constraints on phase error are more restrictive in the K band, and might not be met. This would add to the fact that stellar leaks are already the main trouble in the near-infrared range. Note that PRIMA will not be available before 2003. Another fringe sensor unit (FINITO) will be available at the end of 2001, with lower performances.

- *Wavefront aberrations.* Corrugations in the stellar wavefront caused by atmospheric turbulence and imperfect optics reduce the interferometer fringe visibility and hence the null depth. Adaptive optics will provide low order corrections to the wavefront (MACAO[9] for the VLTI, available for all UTs by early 2004), while single-mode spatial filtering can be used to significantly improve the correction (especially for high order aberrations). The current baseline for MACAO is optimized for use in the 1 - 2.5 μm wavelength range. A high Strehl ratio in the N band would be very beneficial to the nulling instrument. Adaptive optics systems are not scheduled yet for the Auxiliary Telescopes.

- *Scintillation.* Unequal intensities between the two beams of the interferometer produce imperfect fringe visibility. In order to have good nulling performances, an amplitude matching system might be required.

[8]Phase-Referenced Imaging and Microarcsecond Astrometry.

[9]The MACAO-VLTI system is a 60 element curvature adaptive optics system operating at the Coudé focus of one of the VLT UTs delivering a corrected wavefront to the VLTI delay lines.

All these limitations tamper the SNR by an amount that is still to be evaluated in the light of VLTI's instruments performances. Moreover, even if signal-to-noise calculations show that exo-zodi are theoretically detectable, the derivation of their physical parameters from the observations has not been considered yet.

5.4.3 Current projects

Nulling interferometry has already been used with success at 10 μm with two 1.8 m telescopes of the Multiple Mirror Telescope (Hinz *et al.* [22], [21]) without adaptive optics nor fringe tracker, and without cooled optics. When the atmosphere caused the wavefronts to be (by chance) exactly out of phase, the entire stellar Airy pattern disappeared. By selecting the minimal flux images from a series of short exposures, a rejection rate of order 25 has been reached, allowing the detection of dust nebulae around α Orionis and R Leonis. (This was one of the last observations with the MMT, which has now been dismantled and has being replaced by a 6.5 m mirror).

A new nulling instrument (BLINC) which will use two sections of the 6.5 m MMT is currently being built by Hinz *et al.* [23]. It will benefit from an adaptive optics system and a common-path phase-sensing system to suppress all but 10^{-4} of the stellar flux. This level of cancellation, combined with cryogenically-cooled optics, will allow to detect zodiacal emission down to the mJy level at 11 μm, i.e., as faint as 10-20 times the solar level for a star at 10 pc. Two other nulling instruments are scheduled for the Keck (2001) and LBT (2004) interferometers, both with the goal to detect exo-zodiacal clouds at the solar level for a 10 pc star. Note that these two instruments are restricted to the northern hemisphere, so that the VLTI is the only possibility to observe southern hemisphere targets as a precursor to Darwin.

Conclusions

Since its 1993 proposal, the IRSI-Darwin concept, aimed at exoplanet detection by means of infrared interferometry, has evolved in many ways: on the one hand, studies have proven the theoretical feasibility of the mission, and suggested ways to improve its efficiency; on the other hand, technological developments are currently underway, on ground-based instruments as well as in the context of space-borne missions, in order to demonstrate the practical feasibility of Darwin.

The present work is a theoretical contribution to the development of the Darwin nulling interferometer. Nulling interferometry is a technique intended to suppress all the light coming from a blinding star by means of destructive interference, in order to reveal the star's potential planetary companions. Once the starlight has been cancelled, the main obstacle to planet detection is the emission of zodiacal dust in our own solar system as well as around the target star. The present thesis mostly focuses on internal modulation, a recent technique aimed at cancelling such spurious background signals in a single process.

The basic idea of internal modulation is to apply a variable phase shift between the nulled outputs of two or more interferometers sharing the same telescopes, and then recombine these outputs to make the detection. Under certain conditions, only the signal originating from the planet will be modulated as a result of this process. This signal then can be retrieved easily by synchronous demodulation. In the present work, a comprehensive investigation of possible configurations of a telescope array with internal modulation has been undertaken. This investigation has led to a number of new possible configurations, whose performances have been compared with that of the current Darwin configuration. A factor of two has been gained in the sensitivity of the interferometer to the planetary signal. This improvement will reduce by a factor of four the time required to detect and characterize exoplanets, and therefore will allow the study of many more planetary systems than initially anticipated.

The road-map of theoretical and technological developments leading to the final Darwin design is still very long, so that the spacecraft is not likely to be launched before 2012. In the context of selecting a configuration, the present work should be followed by two kinds of studies. The first one concerns the synthesis of an image with a nulling interferometer which, just like any interferometer, only produces fringes but no direct image. The ability of each configuration to recover an image of the planetary system will probably constitute a very important criterion in the choice of the final configuration. The second idea to pursue is the design of a nulling instrument for the VLTI. Signal-to-noise calculations have proven that this instrument should be able to characterize exo-zodiacal clouds down to the 10-zodi level, but much work still remains to be done.

Appendix A

Transmission map symmetry

After reading section 3.2.3, one may think intuitively that a real generalized pupil is a necessary and sufficient condition to have a centro-symmetric transmission map, i.e., an even interferometer response: $R(-\vec{\theta}) = R(\vec{\theta})$. Remember that the generalized entrance pupil (3.6) is real when all phase shifts ϕ_k are whole multiples of π, so that all $e^{j\phi_k}$ are real[1]. We will prove below that this condition is sufficient but not necessary, since non-real pupils can also lead to an even interferometer response.

A.1 Sufficient condition

It is straightforward to prove that $(\phi_k = 0[\pi]) \Rightarrow (R(-\vec{\theta}) = R(\vec{\theta}))$. Let us define the complex quantity $\mathbf{Z}(\vec{\theta})$ by the relation

$$R(\vec{\theta}) = \left| \sum_{k=1}^{n} D_k e^{j2\pi(\vec{L_k}\cdot\vec{\theta})/\lambda} \, e^{j\phi_k} \right|^2 = \left| \mathbf{Z}(\vec{\theta}) \right|^2 . \tag{A.1}$$

The condition $\phi_k = 0[\pi]$ implies that all $e^{j\phi_k}$ are real, so that the only complex factor in $\mathbf{Z}(\vec{\theta})$ is $e^{j2\pi(\vec{L_k}\cdot\vec{\theta})/\lambda}$, which implies

$$\overline{\mathbf{Z}}(\vec{\theta}) = \sum_{k=1}^{n} D_k e^{-j2\pi(\vec{L_k}\cdot\vec{\theta})/\lambda} \, e^{j\phi_k} = \mathbf{Z}(-\vec{\theta}) \, , \tag{A.2}$$

where the "overline" sign represents complex conjugation. Now, it is obvious that

$$R(-\vec{\theta}) = \left| \mathbf{Z}(-\vec{\theta}) \right|^2 = \left| \overline{\mathbf{Z}}(\vec{\theta}) \right|^2 = \left| \mathbf{Z}(\vec{\theta}) \right|^2 = R(\vec{\theta}) \, . \tag{A.3}$$

A.2 Necessary condition

The necessary condition writes $(R(-\vec{\theta}) = R(\vec{\theta})) \Rightarrow (\phi_k = 0[\pi])$. We will prove that this condition is not true when the number of telescopes is larger than two, since in that case

[1] This will be noted $\phi_k = 0[\pi]$ in the following discussion, which means that ϕ_k is equal to zero to within an integer multiple of π.

we may find even transmission maps associated with non-real generalized pupils (i.e., there is at least one k such that $\phi_k \neq 0[\pi]$).

Let us write the equation $R(\vec{\theta}) = R(-\vec{\theta})$ explicitly:

$$
\begin{aligned}
& \left|\mathbf{Z}(\vec{\theta})\right|^2 = \left|\mathbf{Z}(-\vec{\theta})\right|^2 \\
\Rightarrow\ & \mathbf{Z}(\vec{\theta})\overline{\mathbf{Z}}(\vec{\theta}) = \mathbf{Z}(-\vec{\theta})\overline{\mathbf{Z}}(-\vec{\theta}) \\
\Rightarrow\ & \sum_{k=1}^{n} D_k\, e^{j\frac{2\pi}{\lambda}\vec{L_k}\cdot\vec{\theta}}\, e^{j\phi_k} \sum_{l=1}^{n} D_l\, e^{-j\frac{2\pi}{\lambda}\vec{L_l}\cdot\vec{\theta}}\, e^{-j\phi_l} = \sum_{k=1}^{n} D_k\, e^{-j\frac{2\pi}{\lambda}\vec{L_k}\cdot\vec{\theta}}\, e^{j\phi_k} \sum_{l=1}^{n} D_l\, e^{j\frac{2\pi}{\lambda}\vec{L_l}\cdot\vec{\theta}}\, e^{-j\phi_l} \\
\Rightarrow\ & \sum_{k=1}^{n}\sum_{l=1}^{n} D_k D_l\, e^{j\frac{2\pi}{\lambda}(\vec{L_k}-\vec{L_l})\cdot\vec{\theta}}\, e^{j(\phi_k-\phi_l)} = \sum_{k=1}^{n}\sum_{l=1}^{n} D_k D_l\, e^{j\frac{2\pi}{\lambda}(\vec{L_l}-\vec{L_k})\cdot\vec{\theta}}\, e^{j(\phi_k-\phi_l)} \\
\Rightarrow\ & \sum_{k=1}^{n}\sum_{l=1}^{n} D_k D_l\, \sin\left(\tfrac{2\pi}{\lambda}(\vec{L_k}-\vec{L_l})\cdot\vec{\theta}\right)\, e^{j(\phi_k-\phi_l)} = 0
\end{aligned}
$$

The real and imaginary parts of the last equation must cancel out separately for all possible values of $\vec{\theta}$, giving the two following equations:

$$
\left.
\begin{aligned}
&\sum_{k=1}^{n}\sum_{l=1}^{n} D_k D_l\, \sin\left(\tfrac{2\pi}{\lambda}(\vec{L_k}-\vec{L_l})\cdot\vec{\theta}\right)\, \cos(\phi_k-\phi_l) = 0 \\
&\sum_{k=1}^{n}\sum_{l=1}^{n} D_k D_l\, \sin\left(\tfrac{2\pi}{\lambda}(\vec{L_k}-\vec{L_l})\cdot\vec{\theta}\right)\, \sin(\phi_k-\phi_l) = 0
\end{aligned}
\right\} \text{ for all } \vec{\theta}\,. \tag{A.4}
$$

The first of these two equations will always be true, since cosine is even and sine odd: the (k,l) and (l,k) terms are of opposite sign, and will cancel out. On the other hand, the terms with $k = l$ are all zero because the difference $\vec{L_k} - \vec{L_l}$ cancels out. Therefore, the only equation to satisfy is

$$
\sum_{k=1}^{n}\sum_{l=1}^{n} D_k D_l\, \sin\left(\frac{2\pi}{\lambda}(\vec{L_k}-\vec{L_l})\cdot\vec{\theta}\right)\, \sin(\phi_k-\phi_l) = 0 \text{ , for all } \vec{\theta}\,. \tag{A.5}
$$

A.2.1 Two telescopes

If we take the origin of the axes at the middle of the segment joining the two telescopes, then $\vec{L_1} = -\vec{L_2}$, so that (A.5) becomes:

$$
D_1 D_2\, \sin(\frac{2\pi}{\lambda} 2\vec{L_1}\cdot\vec{\theta})\, \sin(\phi_1-\phi_2) + D_2 D_1\, \sin(-\frac{2\pi}{\lambda} 2\vec{L_1}\cdot\vec{\theta})\, \sin(\phi_2-\phi_1) = 0 \ ,\ \forall\,\vec{\theta}\,. \tag{A.6}
$$

This equation requires that $\sin(\phi_1 - \phi_2) = 0$, and so $\phi_2 = \phi_1[\pi]$, which gives a real pupil since we always take $\phi_1 = 0$ as a reference. Thus, in the case of two telescopes, the one and only way to get a centro-symmetric transmission map is to impose a real generalized pupil.

A.2.2 Three telescopes

As in the case of two telescopes, equation (A.5) can be satisfied by cancelling out all the $\sin(\phi_k - \phi_l)$ factors, which leads to $\phi_2 = \phi_1[\pi]$ and $\phi_3 = \phi_1[\pi]$. Once again, choosing $\phi_1 = 0$, we get a real entrance pupil.

But another way to satisfy this equation is to find particular values of $\vec{L_k}$, which would not require all the $\sin(\phi_k - \phi_l)$ factors to vanish. Those values of $\vec{L_k}$ are chosen in order to create a linear relation between two of the $\sin(\frac{2\pi}{\lambda}(\vec{L_k} - \vec{L_l}) \cdot \vec{\theta})$ factors, which will be the case if two of the $\vec{L_k} - \vec{L_l}$ are co-linear. In that case, the three telescopes lie on a single line, and we can take the origin of the axes on that line. Choosing for example $\vec{L_2} = 0$ and $\vec{L_1} = -\vec{L_3} = \vec{L}$ (DAC-like configuration), the condition becomes

$$\begin{aligned} & 2D_1D_2 \sin(\frac{2\pi}{\lambda}\vec{L} \cdot \vec{\theta}) \sin(\phi_1 - \phi_2) + 2D_1D_3 \sin(\frac{2\pi}{\lambda}2\vec{L} \cdot \vec{\theta}) \sin(\phi_1 - \phi_3) \\ & +2D_2D_3 \sin(\frac{2\pi}{\lambda}\vec{L} \cdot \vec{\theta}) \sin(\phi_2 - \phi_3) = 0 \ , \ \forall \vec{\theta} \\ \Rightarrow \quad & D_1D_2 \sin(\phi_1 - \phi_2) + D_2D_3 \sin(\phi_2 - \phi_3) + 2D_1D_3 \cos(\frac{2\pi}{\lambda}\vec{L} \cdot \vec{\theta}) \sin(\phi_1 - \phi_3) = 0 \ , \ \forall \vec{\theta} \ . \end{aligned}$$

The last equation can be satisfied with $\phi_1 = \phi_3[\pi]$ and

$$D_1 \sin(\phi_1 - \phi_2) + D_3 \sin(\phi_2 - \phi_3) = 0 \ .$$

If $\phi_1 = \phi_3 + \pi[2\pi]$, a possible solution is simply $D_1 = D_3$.

So, all the linear configurations with telescopes 1 and 3 at equal distances from telescope 2, satisfying the relations $D_1 = D_3$, $\phi_1 = 0$, $\phi_3 = \pi$, and with arbitrary D_2 and ϕ_2 also give a centro-symmetric transmission map. Since ϕ_2 can take an arbitrary value, we can choose it in order to get a complex entrance pupil, proving that complex pupils may also lead to centro-symmetric maps (i.e., $R(-\vec{\theta}) = R(\vec{\theta})$). The necessary condition is therefore not verified for the case of three telescopes.

A.2.3 More than three telescopes

When there are more than three telescopes, we can always find particular values of $\vec{L_k}$ ($k = 1, 2, \ldots, n$) such that the condition (A.5) does not absolutely require all the $\sin(\phi_k - \phi_l)$ to vanish. Using that trick, complex pupils giving centro-symmetric maps can once again be found. We will not write down the long mathematical developments which lead to those special configurations. Nevertheless, let us take an example with four telescopes: the combination of two Bracewell interferometers in a square configuration with an arbitrary phase shift between the two Bracewell is one of those complex configurations that lead to centro-symmetric maps.

A.3 Consequences

We have just proven that real entrance pupils always yield centro-symmetric maps, whereas complex pupils generally yield asymmetric maps, but sometimes, in particular cases, centro-symmetric maps. What are the consequences of these results?

- When we don't use internal modulation, an asymmetric map is needed to distinguish the planetary signal versus exozodiacal light (section 3.2.3). A complex entrance pupil is therefore needed, and of course it cannot be one of the special cases with a symmetric map.
- When internal modulation is used, it has been proven in section 3.2.5 that real entrance pupils are needed for each of the sub-arrays, which will therefore have a centro-symmetric transmission map.

The conclusion of these remarks is that the complex pupils with a centro-symmetric map are of no interest, and should not be further investigated. On the other hand, both real pupils and complex pupils with asymmetric maps have to be studied in detail.

Appendix B

Circular arrays with a θ^{-4} rejection

A systematic study of circular configurations with a θ^{-4} rejection ratio (or better) is presented below. Circular arrays play an important role for the DARWIN mission, because they do not require long delay lines. As explained in section 3.1.2, a θ^4 starlight suppression is needed to compensate for the huge contrast between the star and the planet.

B.1 Complex entrance pupils

In this section, we will assume that exoplanet detection is performed with a rotating array (no internal modulation). We will explore the possibility to have an asymmetric transmission map (which requires a complex entrance pupil, as proven in appendix A) with a given number of telescopes distributed on a circle, and giving a θ^{-4} rejection rate. The necessity of an asymmetric map is proven in section 3.2.3, which is relevant if internal modulation is not used. We will begin with 3-telescope arrays, since two telescopes are not enough to achieve a θ^4 starlight suppression.

B.1.1 Three telescopes

In appendix C.1, dedicated to linear arrays, we prove that the only 3-telescope configuration giving a θ^4 starlight suppression is the DAC, which has a centro-symmetric transmission map since its entrance pupil is real. So, three telescopes are not enough to allow secure detection of Earth-like exoplanets versus exozodiacal light.

B.1.2 Four telescopes

Let us rewrite the conditions (3.8) for a θ^{-4} rejection rate, assuming $L_k = L$ for all k:

$$\sum_{k=1}^{4} D_k \, e^{j\phi_k} = 0$$

$$\sum_{k=1}^{4} D_k \, \cos(\delta_k - \phi) \, e^{j\phi_k} = 0 \, , \text{ for all } \phi.$$

We will take telescope 1 as a reference for the diameters ($D_1 = 1$), the angular distances ($\delta_1 = 0$) and the phase shifts ($\phi_1 = 0$), so that the above equations write:

$$\begin{aligned}
&D_2\, e^{j\phi_2} + D_3\, e^{j\phi_3} + D_4\, e^{j\phi_4} = -1 \ , \\
&D_2\, \cos\delta_2\, e^{j\phi_2} + D_3\, \cos\delta_3\, e^{j\phi_3} + D_4\, \cos\delta_4\, e^{j\phi_4} = -1 \ , \\
&D_2\, \sin\delta_2\, e^{j\phi_2} + D_3\, \sin\delta_3\, e^{j\phi_3} + D_4\, \sin\delta_4\, e^{j\phi_4} = 0 \ .
\end{aligned}$$

Taking the real and imaginary parts of these equations, and letting $c_k = D_k \cos(\phi_k)$ and $s_k = D_k \sin(\phi_k)$, we obtain a system of six equations with nine unknown parameters (c_k, s_k and δ_k for $k = 2, 3, 4$):

$$\begin{aligned}
&c_2 + c_3 + c_4 = -1 \ , \\
&s_2 + s_3 + s_4 = 0 \ , \\
&c_2 \cos\delta_2 + c_3 \cos\delta_3 + c_4 \cos\delta_4 = -1 \ , \\
&s_2 \cos\delta_2 + s_3 \cos\delta_3 + s_4 \cos\delta_4 = 0 \ , \\
&c_2 \sin\delta_2 + c_3 \sin\delta_3 + c_4 \sin\delta_4 = 0 \ , \\
&s_2 \sin\delta_2 + s_3 \sin\delta_3 + s_4 \sin\delta_4 = 0 \ .
\end{aligned} \tag{B.1}$$

Now, if the positions δ_k of the telescopes are fixed, we get two linear systems of three equations, each with three unknowns, because the equations are decoupled in c_k and s_k. These two systems can be written in a matrix form:

$$\begin{pmatrix} 1 & 1 & 1 \\ \cos\delta_2 & \cos\delta_3 & \cos\delta_4 \\ \sin\delta_2 & \sin\delta_3 & \sin\delta_4 \end{pmatrix} \begin{pmatrix} c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix} \ , \tag{B.2}$$

$$\begin{pmatrix} 1 & 1 & 1 \\ \cos\delta_2 & \cos\delta_3 & \cos\delta_4 \\ \sin\delta_2 & \sin\delta_3 & \sin\delta_4 \end{pmatrix} \begin{pmatrix} s_2 \\ s_3 \\ s_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \ . \tag{B.3}$$

Defining M, the common matrix of these two linear systems, we have two cases to examine, depending on whether $\det M$ is zero or not.

Case $\det M \neq 0$

In that case, there exists one and only one solution for each of the two linear systems. The second system immediately gives the trivial solution $s_k = 0$ for all k, so that $\sin\phi_k = 0$ for all k. This implies that the phase shifts are restricted to $0[\pi]$, and so, that the generalized entrance pupil is real (giving the GAC configurations). Since the phase shifts are restricted to $0[\pi]$, asymmetric transmission maps cannot exist under these conditions.

Case $\det M = 0$

In that case, the three equations of each linear system are not linearly independent any more, so that an infinity of solutions may be found for both systems: c_4 and s_4 can be taken as free parameters, and each value of those parameters will produce a unique solution.

Asymmetric transmission maps with four telescopes on a circle and a θ^4 on-axis suppression could therefore be possible.

But unfortunately, $\det M = 0$ does not correspond to any physical solution with four telescopes. Developing the determinant gives successively:

$$\begin{aligned}
& \cos\delta_3\sin\delta_4 - \sin\delta_3\cos\delta_4 + \cos\delta_4\sin\delta_2 - \sin\delta_4\cos\delta_2 + \cos\delta_2\sin\delta_3 - \sin\delta_2\cos\delta_3 = 0 \\
\Rightarrow\ & \sin(\delta_4-\delta_3) + \sin(\delta_2-\delta_4) + \sin(\delta_3-\delta_2) = 0 \\
\Rightarrow\ & 2\sin(\frac{\delta_4-\delta_3}{2})\left(\cos(\frac{\delta_4-\delta_3}{2}) - \cos(\delta_2 - \frac{\delta_3+\delta_4}{2})\right) = 0\ ,
\end{aligned}$$

which is true if and only if

$$\begin{aligned}
& \delta_4 = \delta_3[2\pi] : \text{impossible, because telescopes 3 and 4 would merge,} \\
\text{or}\quad & \frac{\delta_4-\delta_3}{2} = \delta_2 - \frac{\delta_3+\delta_4}{2}\,[2\pi] \Rightarrow \delta_4 = \delta_2[2\pi] : \text{impossible (same reason),} \\
\text{or}\quad & \frac{\delta_3-\delta_4}{2} = \delta_2 - \frac{\delta_3+\delta_4}{2}\,[2\pi] \Rightarrow \delta_3 = \delta_2[2\pi] : \text{impossible (same reason).}
\end{aligned}$$

Conclusion

We have proven that configurations with four telescopes on a circle and a complex entrance pupil cannot give a θ^{-4} rejection rate. The only circular 4-telescope configurations with such a rejection rate are the GACs (with a real entrance pupil by definition), which have a centro-symmetric transmission map, and so are not appropriate for a distinction between a planet and exozodiacal light.

Case of two merged telescopes

The previous conclusion does not consider one particular case, where two telescopes merge. This is not equivalent to the 3-telescope case since two different phase shifts can be applied to a single telescope (which would be composed of the two merged telescopes). Let us see if a new kind of nulling configuration could be found in that case. We will take the case $\delta_3 = \delta_4$ as an example.

With $\delta_3 = \delta_4$, the system (B.3) can be written as

$$\begin{aligned}
& s_2 = -s_3 - s_4\ , \\
& (\cos\delta_3 - \cos\delta_2)s_3 + (\cos\delta_3 - \cos\delta_2)s_4 = 0\ , \\
& (\sin\delta_3 - \sin\delta_2)s_3 + (\sin\delta_3 - \sin\delta_2)s_4 = 0\ .
\end{aligned} \tag{B.4}$$

Assuming $\delta_2 \neq \delta_3$, this system reduces to $s_3 + s_4 = 0$ and $s_2 = 0$, so that $\phi_2 = 0[\pi]$. Then, the system (B.2) is rewritten with $\delta_3 = \delta_4$:

$$\begin{aligned}
& c_2 = -1 - c_3 - c_4\ , \\
& (\cos\delta_3 - \cos\delta_2)c_3 + (\cos\delta_3 - \cos\delta_2)c_4 = -1 + \cos\delta_2\ , \\
& (\sin\delta_3 - \sin\delta_2)c_3 + (\sin\delta_3 - \sin\delta_2)c_4 = \sin\delta_2\ .
\end{aligned} \tag{B.5}$$

The last two equations of this system cannot be simultaneously verified unless

$$\frac{-1+\cos\delta_2}{\cos\delta_3-\cos\delta_2}=\frac{\sin\delta_2}{\sin\delta_3-\sin\delta_2}\ . \tag{B.6}$$

By means of some trigonometry relations, this equation easily gives the following condition:

$$\tan\left(\frac{\delta_2+\delta_3}{2}\right)=\tan\left(\frac{\delta_2}{2}\right)\ , \tag{B.7}$$

which is satisfied if and only if $\delta_3 = 0[2\pi]$. With this result, the above system of equation yields the condition $c_3+c_4=-1$, so that $c_2=0$. Since we have already proven that $s_2=0$, the only possible solution is $D_2=0$, and telescope 2 vanishes. Thus, there remains only one telescope at $\delta_1=\delta_3=\delta_4=0$, whose output is divided into three beams, which interfere destructively at recombination. This is not an interferometer any more! Moreover, it can be proven that all sources are cancelled by this process: the off-axis planet as well as the on-axis star. Note that the case where $\delta_2=\delta_3$, which has not been considered, would not give any other solution.

This kind of "single aperture interferometer" is used with the Achromatic Interfero Coronagraph (see section 2.3.2), where two beams coming from a single telescope are phase-shifted and rotated before recombination. The 180° rotation of one beam, achieved by a cat's eye system, produces two symmetrical images of an off-axis source, while the on-axis star is perfectly cancelled.

B.1.3 Five telescopes

We prove below that, if the angular positions of the five telescopes on a circle are given, we may find a "double infinity" of solutions with a complex pupil and a θ^{-4} rejection rate, which means that there remains two free parameters when the positions are fixed.

Let us rewrite the set of equations (B.1) in the case of five telescopes on a circle, with the same conventions ($D_1=1$, $\delta_1=0$, $\phi_1=0$):

$$\begin{aligned}
&c_2+c_3+c_4+c_5=-1\ ,\\
&s_2+s_3+s_4+s_5=0\ ,\\
&c_2\cos\delta_2+c_3\cos\delta_3+c_4\cos\delta_4+c_5\cos\delta_5=-1\ ,\\
&s_2\cos\delta_2+s_3\cos\delta_3+s_4\cos\delta_4+s_5\cos\delta_5=0\ ,\\
&c_2\sin\delta_2+c_3\sin\delta_3+c_4\sin\delta_4+c_5\sin\delta_5=0\ ,\\
&s_2\sin\delta_2+s_3\sin\delta_3+s_4\sin\delta_4+s_5\sin\delta_5=0\ .
\end{aligned} \tag{B.8}$$

If the positions δ_k of the telescopes are fixed, we get a set of 6 linear equations with 8 unknown parameters (c_k and s_k for $k=2,3,4,5$). We can take c_5 and s_5 as free parameters, and write as before:

$$\begin{pmatrix}1&1&1\\ \cos\delta_2&\cos\delta_3&\cos\delta_4\\ \sin\delta_2&\sin\delta_3&\sin\delta_4\end{pmatrix}\begin{pmatrix}c_2\\c_3\\c_4\end{pmatrix}=\begin{pmatrix}-1-c_5\\-1-c_5\cos\delta_5\\-c_5\sin\delta_5\end{pmatrix}\ , \tag{B.9}$$

$$\begin{pmatrix}1&1&1\\ \cos\delta_2&\cos\delta_3&\cos\delta_4\\ \sin\delta_2&\sin\delta_3&\sin\delta_4\end{pmatrix}\begin{pmatrix}s_2\\s_3\\s_4\end{pmatrix}=\begin{pmatrix}-s_5\\-s_5\cos\delta_5\\-s_5\sin\delta_5\end{pmatrix}\ . \tag{B.10}$$

We have already proved that $\det M = 0$ does not have any physical meaning, and so we can restrict ourselves to the case $\det M \neq 0$. In that case, for each value of c_5 and s_5, i.e., of D_5 and ϕ_5, there exists one and only one solution for each of the two linear systems, giving a unique solution for the diameters and phase shifts of the three remaining telescopes. Consequently, asymmetric transmission maps can be obtained with five telescopes on a circle, and with a θ^{-4} rejection rate. Five telescopes arrays are therefore well suited for exoplanet detection with a good rejection of exozodiacal light (once again, we do not consider internal modulation here: the only modulation comes from the rotation of the array).

B.2 Real entrance pupils

Let us now consider the case of real generalized pupils. This kind of pupils is used with internal modulation, where they are recombined pairwise with a variable phase shift.

B.2.1 Three telescopes

Configurations with three telescopes on a circle cannot result in a θ^{-4} rejection rate, as proven in appendix C.

B.2.2 Four telescopes

Using the developments of paragraph B.1.2, we can prove an interesting result concerning the GACs, which correspond to the case $\det M \neq 0$ as discussed above. We have seen that, under this condition, a unique solution exists for the $c_k = D_k \cos\phi_k$ and that the phase shifts are restricted to $\phi_k = 0[\pi]$, which gives a real entrance pupil. So, for each set of δ_k (position of the telescopes), there exists one and only one GAC configuration (real circular configuration with a θ^{-4} rejection rate), defined by the values of D_k and ϕ_k ($\phi_k = 0$ if c_k is positive, $\phi_k = \pi$ if c_k is negative).

B.2.3 Five telescopes

We can use the results obtained in the complex pupil case, with $s_k = 0$ since all ϕ_k are restricted to integer multiples of π. Taking c_5 as a free parameter, the linear system B.9 remains valid:

$$\begin{pmatrix} 1 & 1 & 1 \\ \cos\delta_2 & \cos\delta_3 & \cos\delta_4 \\ \sin\delta_2 & \sin\delta_3 & \sin\delta_4 \end{pmatrix} \begin{pmatrix} c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} -1 - c_5 \\ -1 - c_5\cos\delta_5 \\ -c_5\sin\delta_5 \end{pmatrix} . \tag{B.11}$$

Since $\det M \neq 0$, there exists a unique solution for each set of δ_k and for each c_5. So, for each position of the telescopes, an infinity of real configurations can achieve the θ^{-4} rejection rate.

Now, a new question arises: is it possible to get a θ^6 on-axis extinction with five telescopes on a circle? Unfortunately, the answer is negative, as proven below. Let us write the additional condition for a θ^{-6} rejection:

$$\sum_{k=1}^{5} D_k \cos^2(\delta_k - \phi)\, e^{j\phi_k} = 0\,, \text{ for all } \phi. \tag{B.12}$$

Using some well-known trigonometry formulae, this equation leads to

$$\sum_{k=1}^{5} c_k \cos(2\delta_k) = 0 \text{ , and } \sum_{k=1}^{5} c_k \sin(2\delta_k) = 0\,, \tag{B.13}$$

giving a new system of five equations with five unknowns (in order to get a square matrix, we will not impose $c_1 = 1$ this time):

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ \cos\delta_1 & \cos\delta_2 & \cos\delta_3 & \cos\delta_4 & \cos\delta_5 \\ \sin\delta_1 & \sin\delta_2 & \sin\delta_3 & \sin\delta_4 & \sin\delta_5 \\ \cos 2\delta_1 & \cos 2\delta_2 & \cos 2\delta_3 & \cos 2\delta_4 & \cos 2\delta_5 \\ \sin 2\delta_1 & \sin 2\delta_2 & \sin 2\delta_3 & \sin 2\delta_4 & \sin 2\delta_5 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}\,. \tag{B.14}$$

Defining M as the square matrix of this system, we immediately see that $\det M \neq 0$ gives the unique solution $c_k = 0$ for all k, which implies that $D_k = 0$ (impossible). So, let us see the conditions to have $\det M = 0$. Taking $\delta_1 = 0$ as a reference, the determinant writes:

$$\begin{aligned} \det M \;=\; & 256 \sin(\tfrac{\delta_2}{2}) \sin(\tfrac{\delta_3}{2}) \sin(\tfrac{\delta_4}{2}) \sin(\tfrac{\delta_5}{2}) \sin(\tfrac{\delta_2-\delta_3}{2}) \sin(\tfrac{\delta_2-\delta_4}{2}) \\ & \sin(\tfrac{\delta_2-\delta_5}{2}) \sin(\tfrac{\delta_3-\delta_4}{2}) \sin(\tfrac{\delta_3-\delta_5}{2}) \sin(\tfrac{\delta_4-\delta_5}{2})\,. \end{aligned}$$

For this expression to vanish, two of the δ_k must be equal to within a whole multiple of 2π (remember that $\delta_1 = 0$); so that two telescopes merge, which is not possible. Therefore, a circular configuration with five telescopes cannot give a θ^{-6} rejection rate. This holds true for complex pupils, since then we are faced with the same matrix M in the complex case. The only difference is the presence of two decoupled systems, one for the c_k and the other for the s_k.

Appendix C

Linear arrays with a θ^{-4} rejection

This time, we will not separate the developments for complex and real pupils any more. A general study of linear arrays is done with arbitrary phase shifts, and real pupils are considered as a special case of complex ones. Once again, we will restrict the study to configuration with a θ^{-4} rejection rate or better.

Linear configurations are particular because long delays lines can generally be avoided, if the distances between the telescopes are equal.

C.1 Three telescopes

It this first section, we will prove that the DAC family (Degenerate Angel Cross, linear array defined in section 3.2.2)) is the only three-telescope configuration giving a θ^{-4} rejection rate, by showing that non-linear configurations cannot result in a θ^{-4} rejection rate, and that the only possible linear array is a DAC.

Consider three telescopes in any triangular shape (non-linear configuration). We can always draw a circle defined by the centers of the three telescopes (figure C.1).

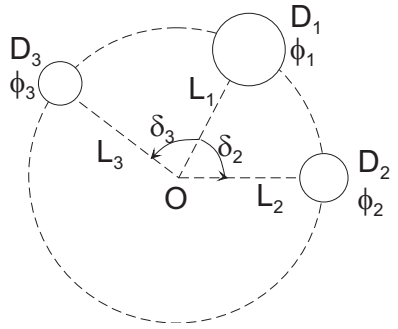


Figure C.1: Three non-aligned telescopes can always be inscribed on a circle (center O).

Taking the center O of the circle as the origin of the frame, and the direction of telescope number 1 as an angular reference (so that $\delta_1 = 0$), we write the conditions (3.8) for a θ^{-4}

rejection rate:

$$\sum_{k=1}^{3} D_k\, e^{j\phi_k} = 0\ ,$$
$$\sum_{k=1}^{3} D_k\, L_k\, \cos(\delta_k - \phi)\, e^{j\phi_k} = 0\ ,\ \text{for all } \phi.$$

In a circular configuration, all the telescopes are at equal distances from the origin ($L_k = L$ for all k). If we develop the cosines in the second condition and use the fact that this equation must be true for all ϕ, we get a set a three conditions to satisfy:

$$D_1\, e^{j\phi_1} + D_2\, e^{j\phi_2} + D_3\, e^{j\phi_3} = 0\ , \tag{C.1}$$
$$D_1\, e^{j\phi_1} + D_2 \cos\delta_2\, e^{j\phi_2} + D_3 \cos\delta_3\, e^{j\phi_3} = 0\ , \tag{C.2}$$
$$D_2 \sin\delta_2\, e^{j\phi_2} + D_3 \sin\delta_3\, e^{j\phi_3} = 0\ , \tag{C.3}$$

because $\cos\delta_1 = 1$ and $\sin\delta_1 = 0$.

Then, we take ϕ_2 as the reference for the phase shifts, choosing $\phi_2 = 0$. The equation (C.3) cannot be satisfied unless ϕ_3 is a whole multiple of π, because the second term must also be real. We can always choose the telescope number 1 so that $\sin\delta_2$ and $\sin\delta_3$ are of opposite sign (see figure C.1). It that case, ϕ_3 is restricted to whole multiple of 2π because the two terms must be of different signs to cancel out (D_2 and D_3 are positive quantities). As the last two terms of equation (C.1) are both real and positive, the first term must be a negative real number, giving $\phi_1 = \pi$ to within a whole multiple of 2π. Let us rewrite the first two equations with these values of ϕ_k:

$$-D_1 + D_2 + D_3 = 0\ ,$$
$$-D_1 + D_2 \cos\delta_2 + D_3 \cos\delta_3 = 0\ .$$

Those two equations cannot be simultaneously satisfied unless $\delta_2 = \delta_3 = 0$. Thus, the three telescopes must be at the same place, which is absurd. We conclude that non-linear configurations cannot give a θ^{-4} rejection rate.

The last thing to do is to find the conditions for a linear array to be a DAC. Letting $L_2 = 0$ in the nulling equation, with $\delta_3 = 0$ and $\delta_1 = \pi$ (conditions for a linear array with the origin O of the frame at the center of telescope 2), we obtain:

$$D_1\, e^{j\phi_1} + D_2\, e^{j\phi_2} + D_3\, e^{j\phi_3} = 0\ , \tag{C.4}$$
$$-D_1 L_1\, e^{j\phi_1} + D_3 L_3\, e^{j\phi_3} = 0\ . \tag{C.5}$$

Once again, we choose a reference phase shift $\phi_1 = 0$, so that equation (C.5) gives $\phi_3 = 0$ and $D_1 L_1 = D_3 L_3$. Replacing those values in equation (C.4) yields $\phi_2 = \pi$ and $D_2 = D_1 + D_3$, which is the definition of a DAC. Some example of DACs are proposed in figure C.2. Note that the DACs always have a real entrance pupil.

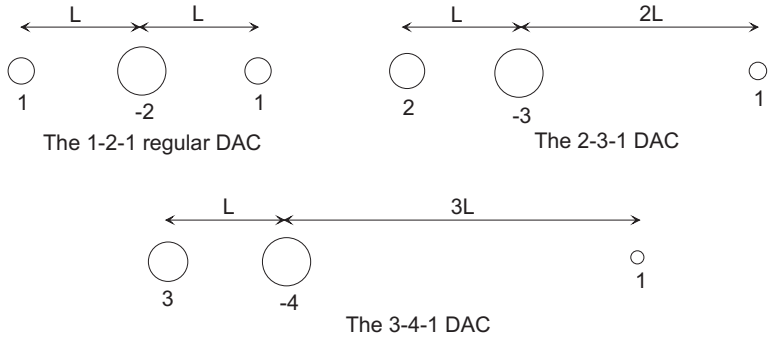


Figure C.2: Some examples of DACs, with their generalized entrance pupil.

C.2 Four telescopes

C.2.1 With a θ^{-4} rejection rate

For the sake of simplicity, we adopt the following conventions (see figure C.3): $L_1 = 0$ (origin of the L axis), and telescope 1 on the left end of the array, so that $\delta_k = 0$ ($\forall k$). The L_k are positive quantities as usual. Under these assumptions, the conditions for a θ^{-4} rejection rate read:

$$D_1\, e^{j\phi_1} + D_2\, e^{j\phi_2} + D_3\, e^{j\phi_3} + D_4\, e^{j\phi_4} = 0\ , \tag{C.6}$$

$$D_2 L_2\, e^{j\phi_2} + D_3 L_3\, e^{j\phi_3} + D_4 L_4\, e^{j\phi_4} = 0\ . \tag{C.7}$$

Since we have many more unknowns than equations, we will take the positions L_k of the telescopes as fixed parameters. The second equation gives a complex relation (i.e., two real relations) between telescopes 2, 3 and 4. The first equation allows us to deduce the characteristics of telescope 1 when the other three telescopes are known.

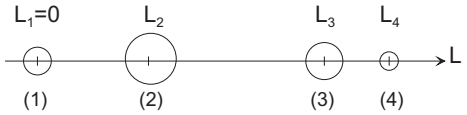


Figure C.3: Conventions for the 4-telescope linear arrays.

Complex configurations

Equation (C.7) splits into two real relations:

$$D_2 L_2 \cos\phi_2 + D_3 L_3 \cos\phi_3 + D_4 L_4 \cos\phi_4 = 0$$
$$D_2 L_2 \sin\phi_2 + D_3 L_3 \sin\phi_3 + D_4 L_4 \sin\phi_4 = 0\ .$$

If the phase shifts ϕ_k ($k = 2, 3, 4$) are fixed, there remains two linear equations with three unknowns (D_2, D_3 and D_4). Since we can always fix $D_2 = 1$ as a reference, there

are in fact only two unknown parameters, which can be deduced by solving the system. Equation (C.6) is then split into two real relations, which easily give D_1 and ϕ_1. So, for a given geometry, there exists a double infinity of linear configurations with a θ^{-4} rejection rate, depending on the relative values of ϕ_3 and ϕ_4 with respect to the $\phi_2 = 0$ reference.

Real configurations

When the ϕ_k are restricted to whole multiples of π, equations (C.6) and (C.7) are reduced to two real relations. Letting $c_k = D_k \cos(\phi_k)$, we get

$$\begin{aligned} c_1 + c_2 + c_3 + c_4 &= 0 \ , \\ c_2 L_2 + c_3 L_3 + c_4 L_4 &= 0 \ . \end{aligned}$$

Taking $c_2 = 1$ as a reference and c_3 as a free parameter, the second equation gives c_4 and then the first gives c_1. Thus, for a given linear geometry, there is an infinity of real configurations with a θ^4 starlight suppression, depending on the value of c_3 with respect to c_2.

C.2.2 With a θ^{-6} rejection rate

In addition to equations (C.6) and (C.7), the following complex relation must also be satisfied (see equation 3.8):

$$D_2 L_2^2 \, e^{j\phi_2} + D_3 L_3^2 \, e^{j\phi_3} + D_4 L_4^2 \, e^{j\phi_4} = 0 \ . \tag{C.8}$$

Letting $c_k = D_k \cos(\phi_k)$ and $s_k = D_k \sin(\phi_k)$, these three equations are equivalent to a system of six real equations:

$$\begin{aligned} c_1 + c_2 + c_3 + c_4 &= 0 \ , \\ s_1 + s_2 + s_3 + s_4 &= 0 \ , \\ c_2 L_2 + c_3 L_3 + c_4 L_4 &= 0 \ , \\ s_2 L_2 + s_3 L_3 + s_4 L_4 &= 0 \ , \\ c_2 L_2^2 + c_3 L_3^2 + c_4 L_4^2 &= 0 \ , \\ s_2 L_2^2 + s_3 L_3^2 + s_4 L_4^2 &= 0 \ . \end{aligned}$$

Now, we can take $\phi_1 = 0$ as a phase reference, so that $s_1 = 0$. In that case, the three equations involving the s_k variables immediately give $s_2 = s_3 = s_4 = 0$, and thus a real entrance pupil. Then, choosing $c_1 = 1$, there remains three equations with three unknowns (c_2, c_3 and c_4), which give a unique solution for the entrance pupil. So, for each given linear geometry, there exits one and only one configuration with a θ^6 on-axis extinction, and this configuration has a real entrance pupil.

A particular subset of this family consists in the superposition of two Bracewell interferometers with different spacings, whose outputs are combined with 180° achromatic phase difference. This configuration has been proposed by Angel and Woolf [3], and constitutes the current baseline for the Terrestrial Planet Finder mission (TPF). This set of

configurations is dubbed OASES (Outpost for Analysis and Spectroscopy of Exo Systems, see [58] for history). Figure 3.19 illustrates the interesting case where the telescopes are at equal distances of each other. In that case, the telescope sizes are in a 1:3:3:1 ratio.

Appendix D

Arbitrary arrays with a θ^{-4} rejection

It is a tedious task to study arbitrary arrays in an analytical way, because one must take into account a very large number of free parameters: the telescopes positions (L_k, δ_k), diameters D_k and phase shifts ϕ_k. Moreover, when more than four telescopes are used, a third order recombination scheme has to be considered, which increases the recombination losses. For those reasons, we will restrict the study to 4-telescope arrays.

Using the same notations as in appendix B and introducing the L_k parameters, the nulling equations read

$$\sum_{k=1}^{4} D_k\, e^{j\phi_k} = 0$$

$$\sum_{k=1}^{4} D_k\, L_k\, \cos(\delta_k - \phi)\, e^{j\phi_k} = 0\,, \text{ for all } \phi.$$

Defining $c_k = D_k \cos\phi_k$ and $s_k = D_k \sin\phi_k$ as usual, and taking telescope 1 as a reference ($L_1 = D_1 = 1$, $\delta_1 = \phi_1 = 0$), the nulling equations can be written in the following matrix form:

$$\begin{pmatrix} 1 & 1 & 1 \\ L_2\cos\delta_2 & L_3\cos\delta_3 & L_4\cos\delta_4 \\ L_2\sin\delta_2 & L_3\sin\delta_3 & L_4\sin\delta_4 \end{pmatrix} \begin{pmatrix} c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix}\,, \tag{D.1}$$

$$\begin{pmatrix} 1 & 1 & 1 \\ L_2\cos\delta_2 & L_3\cos\delta_3 & L_4\cos\delta_4 \\ L_2\sin\delta_2 & L_3\sin\delta_3 & L_4\sin\delta_4 \end{pmatrix} \begin{pmatrix} s_2 \\ s_3 \\ s_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\,. \tag{D.2}$$

As in section B.1.2, if M represents the common matrix of these two linear systems, the case $\det M \neq 0$ leads to a unique real configuration (since all $s_k = 0$, so that all $\phi_k = 0[\pi]$). But what we are looking for is a complex configuration, with which inherent modulation can be used (see section 4.4.4). Developing the condition $\det M = 0$ leads to one relation between the free parameters:

$$L_4 = \frac{L_2 L_3 \sin(\delta_2 - \delta_3)}{L_2\sin(\delta_2 - \delta_4) + L_3\sin(\delta_3 - \delta_4)}\,. \tag{D.3}$$

In that case, the linear system D.2 gives an infinity of solutions for the s_k parameters, depending on the arbitrary value of s_2. On the other hand, the linear system D.1 gives an infinity of solutions if and only if the right-hand vector is a linear combination of the matrix columns. This is the case if the determinant of

$$\begin{pmatrix} 1 & 1 & -1 \\ L_2 \cos\delta_2 & L_3 \cos\delta_3 & -1 \\ L_2 \sin\delta_2 & L_3 \sin\delta_3 & 0 \end{pmatrix} \tag{D.4}$$

vanishes. This condition leads to another relation between the free parameters:

$$L_3 = \frac{L_2 \sin\delta_2}{L_2 \sin(\delta_2 - \delta_3) + \sin\delta_3} \ . \tag{D.5}$$

When this condition is satisfied, the parameters c_3 and c_4 can be expressed as a function of the arbitrary value of c_2.

All in all, only six free parameters remain: the angular positions δ_2, δ_3, δ_4, and the characteristics of telescope 2 (L_2, c_2 and s_2, or equivalently, L_2, D_2 and ϕ_2). These six parameters must be fixed in order to define a unique complex configuration. Note that we can always find two complex configurations with the same geometrical arrangement, by only changing the signs of the phase shifts.

Bibliography

[1] J. Angel, *Ground-based imaging of extrasolar planets using adaptive optics*, Nature **368** (1994), 203.

[2] J. Angel, A. Cheng, and N. Woolf, *A space telescope for infrared spectroscopy of earth-like planets*, Nature **322** (1986), 341.

[3] J. Angel and N. Woolf, *An imaging nulling interferometer to study extrasolar planets*, Nature **322** (1986), 341.

[4] D. Backman, L. Caroff, S. Sandford, and D. Wooden (eds.), *Exozodiacal Dust Workshop*, NASA/CP 1998-10155, 1998.

[5] P. Baudoz, Y. Rabbia, and J. Gay, *Achromatic Interfero Coronagraphy I: Theoretical capabilities for ground-based observations*, Astr. Astrophys. Suppl. Series **141** (2000), 319.

[6] ———, *Study and test of a hybrid coronagraph*, Proc. SPIE **4007** (2000).

[7] P. Baudoz, Y. Rabbia, J. Gay, R. Burg, L. Petro, B. Fleury, P. Madec, and F. Charbonnier, *Achromatic Interfero Coronagraphy II: Effective performance on the sky*, Astr. Astrophys. Suppl. Series **145** (2000), 341.

[8] C. Beichman and T. Velusamy, *Sensitivity of the TPF Interferometer for Planet Detection*, ASP Conf Series **194** (1999), 408.

[9] C. Beichman, N. Woolf, and C. Lindensmith, *The Terrestrial Planet Finder*, Jet Propulsion Laboratory: Pasadena, CA, 1999, JPL 99-3.

[10] D. Bennett, S. Rhie, A. Becker, N. Butler, J. Dann, S. Kapsi, E. Leibowitz, Y. Lipkin, D. Maoz, H. Mendelson, B. Peterson, J. Quinn, O. Shemmer, S. Thomson, and S. Turner, *Gravitational Microlensing Evidence for a Planet Orbiting a Binary Star System*, Nature **402** (1999), 57.

[11] A. Boccaletti, A. Labeyrie, and R. Ragazzoni, *Preliminary results of dark-speckle stellar coronagraphy*, Astron. Astrophys. **338** (1998), 106.

[12] A. Boccaletti, C. Moutou, L. Abe, F. Vakili, A. LAbeyrie, P. Riaud, and J. Schneider, *NEF: NGST Exo-planet Finder: I. Concepts*, ASP Conf. Series **207** (2000), 179.

[13] A. Boccaletti, P. Riand, C. Moutou, and A. Labeyrie, *Snapshot Coronagraphy withan Interferometer in Space*, Icarus **145** (2000), 628.

[14] M. Born and E. Wolf, *Principles of Optics*, 7th (expanded) ed., Cambridge University Press, 1999.

[15] R. Bracewell, *Detecting nonsolar planets by spinning an infrared interferometer*, Nature **274** (1978), 780.

[16] R. Bracewell and R. MacPhie, *Searching for nonsolar planets*, Icarus **38** (1979), 136.

[17] D. Charbonneau, T. Brown, D. Latham, and M. Mayor, *Detection of planetary transit across a Sun-like star*, Astrophys. J. Lett. **529** (2000), 45.

[18] ESA, *DARWIN, the Infrared Space Interferometer*, July 2000, Concept and Feasibility Study Report, ESA-SCI(2000)12.

[19] J. Gay and Y. Rabbia, *Principe d'un coronographe interférentiel*, C. R. Acad. Sci. Paris **322** (1996), Série II b, 265.

[20] O. Guyon and F. Roddier, *Direct exoplanet imaging possibilities of the nulling stellar coronagraph*, Proc. SPIE **4006** (2000), 377, Interferometry in optical astronomy, Munich, Germany.

[21] P. Hinz, R. Angel, W. Hoffmann, D. McCarthy, P. McGuire, M. Cheselka, J. Hora, and N. Woolf, *Imaging Circumstellar Environments with a Nulling Interferometer*, Nature **395** (1998), 251.

[22] ———, *First results of nulling interferometry with the Multiple Mirror Telescope*, Proc. SPIE **3350** (1998), 439.

[23] P. Hinz, R. Angel, N. Woolf, W. Hoffmann, and D. McCarthy, *Imaging Extra-solar Systems from the Ground: The MMT and LBT Nulling Interferometers*, ASP Conf. Series **194** (1999), 401.

[24] A. Karlsson and B. Mennesson, *The Robin Laurance nulling interferometers*, Proc. SPIE **4006** (2000), Interferometry in optical astronomy, Munich, Germany.

[25] A. Labeyrie, *Images of exo-planets obtainable from dark speckls in adaptive telescopes*, Astron. Astrophys. **298** (1995), 544.

[26] ———, *Resolved imaging of extra-solar planets with future 10-100 km optical interferometric arrays*, Astron. Astrophys. Suppl. Ser. **118** (1996), 517.

[27] ———, *Exo-Earth Imager for Exoplanet Snapshots with Resolved Detail*, ASP Conference Series **194** (1999), 350.

[28] P. Lawson (ed.), *Principles of Long Baseline Stellar Interferometry*, Michelson Fellowship Program, 1999.

[29] P. Lawson, P. Dumont, and M. Colavita, *Interferometer Designs for the Terrestrial Planet Finder*, ASP Conference Series **194** (1999), 423.

[30] O. Lay, G. Blackwood, S. Dubovitsky, P. Gorham, and R. Linfield, *Desing for the ST3 Formation Flying Interferometer*, ASP Conference Series **194** (1999), 366.

[31] A. Léger, J.-M. Mariotti, B. Mennesson, M. Ollivier, J.-L. Puget, D. Rouan, and J. Schneider, *Could we search for primitive life on extrasolar planets in the near future? The DARWIN project*, Icarus **123** (1996), 249.

[32] A. Léger, M. Ollivier, K. Altwegg, and N. Woolf, *Is the presence of H_2O and O_3 in an exoplanet a reliable signature of a biological activity?*, Astron. Astrophys. **341** (1999), 304.

[33] P. Léna, *Méthodes physiques de l'observation*, InterÉditions / CNRS Editions, 1996, deuxième édition.

[34] G. Lund, *IRSI-DARWIN*, Tech. report, Alcatel Space, July 2000, Executive Summary, ASPI-00-OSM/IF.21.

[35] F. Malbet, J. Yu, and M. Shao, *High-Dynamic-Range Imaging Using a Deformable Mirror for Space Coronagraphy*, PASP **107** (1995), 386.

[36] M. Mayor and D. Queloz, *A Jupiter-mass companion to solar-type star*, Nature **378** (1995), 355.

[37] B. Mennesson, *Interférométrie stellaire dans l'infrarouge thermique: observations d'environnements circumstellaires par optique guidée monomode et contributions à la mission spatiale DARWIN*, Ph.D. thesis, Université de Paris VII, December 1999.

[38] B. Mennesson, *Further exploration of "Robin Laurance" type configurations for DARWIN*, draft, 2000.

[39] B. Mennesson and A. Léger, *Direct detection and characterization of extrasolar planets: the Mariotti space interferometer*, submitted to Icarus, 1999.

[40] B. Mennesson and J.-M. Mariotti, *Array configuration for a space infrared nulling interferometer dedicated to the search of earthlike extrasolar planets*, Icarus **128** (1997), 202.

[41] B. Mennesson, M. Ollivier, and C. Ruilier, *On the use of single-mode waveguides to correct the optical defects of a nulling interferometer*, submitted to JOSA, 1999.

[42] R. Morgan and J. Burge, *Initial Results of a White Light Nulled Fringe*, ASP Conference Series **194** (1999), 396.

[43] M. Ollivier, *Contribution à la recherche d'exoplanètes - Coronographie interférentielle pour la mission DARWIN*, Ph.D. thesis, Université de Paris XI, December 1999.

[44] M. Ollivier and J. M. Mariotti, *Improvement in the rejection rate of a nulling interferometer by spatial filtering*, Applied Optics **36** (1997), 22, 5340.

[45] F. Paresce, *An interferometry primer*, February 2001, ESO internal publication.

[46] M. Perryman, *Extra-solar planets*, Rep. Prog. Phys. **63** (2000), 1209.

[47] Y. Rabbia, P. Baudoz, and J. Gay, *Achromatic Interfero Coronagraphy & NGST*, The Next Generation Space Telescope: Science Drivers and Technological Challenges, 34th Liège Astrophysics Colloquium, June 1998, p. 279.

[48] Y. Rabbia, J. Gay, and E. Bascou, *Achromatic phase shifter for nulling interferometry*, unpublished, 2001.

[49] F. Roddier and C. Roddier, *Stellar Coronagraph with Phase Mask*, PASP **109** (1997), 815.

[50] D. Rouan, P. Riaud, A. Boccaletti, Y. Clénet, and A. Labeyrie, *The Four-Quadrant Phase-Mask Coronagraph: I. Principle*, PASP **112** (2000), 1479.

[51] E. Serabyn, *Nulling Interferometry: Symmetry Requirements and Experimental Results*, Proc. SPIE **4006** (2000), 437, Interferometry in optical astronomy, Munich, Germany.

[52] E. Serabyn, J. Wallace, H. Nguyen, E. Schmidtlin, and G. Hardy, *Nulling Interferometry: Working on the Dark Fringe*, ASP Conference Series **194** (1999), 437.

[53] J. Surdej, *Astrophysique et Techniques Spatiales*, University of Liège, 2000, course notes.

[54] E. Thomas, *Contributions à l'interférométrie spatiale – Astrométrie globale et détection de planètes extra-solaires*, Ph.D. thesis, Université de Nice-Sophia Antipolis, March 2000.

[55] T. Velusamy, C. Beichman, and M. Shao, *A Dual 3-Element Nulling Interferometer for TPF*, ASP Conference Series **194** (1999), 430.

[56] O. Wallner, K. Kudielka, and W. Leeb, *Nulling Interferometry for Spectroscopic Investigation of Exoplanets - A Statistical Analysis of Imperfections*, Proc. SPIE **4273** (2000), 29.

[57] A. Wolszczan and D. Frail, *A Planetary System Around the Millisecond Pulsar PSR1257+12*, Nature **355** (1992), 145.

[58] N. Woolf, *Planet Finder Options II*, Infrared Space Interferometry: Astrophysics & the Study of Earth-Like Planets (1997), 283, C. Eiroa et al. (eds).

[59] N. Woolf and J. Angel, *Planet Finder Options I: New Linear Nulling Array Configurations*, ASP Conference Series **119** (1997), 285.

[60] ———, *Astronomical searches for earth-like planets and signs of life*, Annu. Rev. Astron. Astrophys. **36** (1998), 507.

[61] N. Woolf, J. Angel, C. Beichman, J. Burge, M. Shao, and D. Tenerelli, *Planet Discoverer Interferometer I: PDI, a potential precursor to Terrestrial Planet Finder*, Proc. SPIE **3350** (1998), 683.